

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 10.2637 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989a-1  
Perfect score: 166  
Sequence: 1 AKKYAKKEKAAYKAYKEAKAKAAEAAKEAYEA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	50.6	384	2 B43592	outer membrane pro
2	80.5	48.5	461	2 T03561	hypothetical prote
3	76	45.8	347	2 E83525	TolA protein PA097
4	74	44.6	372	2 G64064	outer membrane int
5	72	43.4	210	2 A25550	histone H1 - sea u
6	71.5	43.1	211	2 A28100	histone H1-beta, e
7	70.5	42.5	445	2 T50972	probable zootin [i
8	70	42.2	206	2 S09388	histone H1 - sea u
9	70	42.2	277	2 T34625	probable NLP/P60 f
10	70	42.2	1015	2 JC6552	DNA topoisomerase
11	70	42.2	1130	2 T34081	hypothetical prote
12	69.5	41.9	421	2 JY0057	TolA protein - Esc
13	67	40.4	394	2 F90725	membrane spanning
14	67	40.4	394	2 G85576	membrane spanning
15	67	40.4	1021	2 G75403	DNA topoisomerase
16	66.5	40.1	1217	2 A26721	histone H1-gamma,
17	66.5	40.1	1052	1 A44937	kinetoplast-associ
18	66	39.8	231	2 S59599	histone H1 - Chlam
19	66	39.8	318	2 H96535	hypothetical prote
20	66	39.8	1390	2 S51364	sperm tail-specifi
21	65.5	39.5	214	2 G70673	probable hupB - My
22	65.5	39.5	1528	2 Ae0338	surface antigen A
23	65	39.2	101	2 H59099	hypothetical prote
24	64.5	38.9	185	2 A32137	histone H1-delta -
25	64	38.6	241	2 JN0748	histone H1-II - Vo
26	64	38.6	335	2 T21503	hypothetical prote
27	63.5	38.3	621	2 A57591	Id-associated prot
28	63	38.0	248	1 HSUR1P	histone H1, gonada
29	63	38.0	261	2 JN0747	histone H1-I - Vol

hypothetical prote  
hypothetical prote  
probable erythrocy  
histone H1 homolog  
histone H1 (clone  
T06257  
penicillin-binding  
histone H1 - rainb  
histone H1 - midge  
hypothetical prote  
TolA protein [impo  
zootin-like protei  
proton pump intera  
penicillin-binding  
translation initia  
DNA topoisomerase  
hypothetical prote

ALIGNMENTS

RESULT 1

B43592

outer membrane protein TmpB - Treponema phagedenis

C;Species: Treponema phagedenis

C;Date: 30-Jan-1993 #sequence\_revision 12-Mar-1993 #text\_change 08-Oct-1999

C;Accession: B43592

R;Yelton, D.B.; Limberger, R.J.; Curci, K.; Malinosky-Rummell, F.; Slivienski, L.; Schou

Infect. Immun. 59, 3685-3693, 1991

A;Title: Treponema phagedenis encodes and expresses homologs of the Treponema pallidum Tr

A;Reference number: A43592; MUID:91372983; PMID:1894368

A;Accession: B43592

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-384 <YEL>

A;Cross-references: GB:M58563; NID:G155066; PIDN:AAA27480.1; PID:G155067

A;Note: the authors translated the codon TTC for residue 316 as Tyr, and CGA for residue

C;Keywords: membrane protein

Query Match 50.6%; Score 84; DB 2; Length 384;

Best Local Similarity 64.9%; Pred. No. 0.094;

Matches 24; Conservative 5; Mismatches 6; Indels 2; Gaps 2;

Qy 1 AKKYAKKEKAAYKAYKEAKAKAA-EAAAKEAYEA 35

Db 218 AREMAAKEAKAKDKAAKEEAARKAAEEAAAKAAEEA 254

:::|||||:::|||||:::|||||:::|||||:::|||||

AREMAAKEAKAKDKAAKEEAARKAAEEAAAKAAEEA 254

RESULT 2

T03561

hypothetical protein - Rhodobacter capsulatus

C;Species: Rhodobacter capsulatus

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C;Accession: T03561

R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.

Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A;Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003

A;Reference number: Z14955; MUID:97404404; PMID:9256491

A;Accession: T03561

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-461 <VLC>

A;Cross-references: EMBL:AF010496; NID:G3128256; PIDN:AAC16214.1; PID:G3128362

C;Genetics:

A;Map position: 1

Query Match 48.5%; Score 80.5; DB 2; Length 461;

Best Local Similarity 60.0%; Pred. No. 0.25;

Matches 21; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAAYKAYKEAKAKAAEAAAKEAYEA 35

Db 30 AKEVVEKAARKAAKEAKAR-EAAAKDKAAKA 63

:::|||||:::|||||:::|||||:::|||||:::|||||



A;Reference number: JC6552; MUID:98245940; PMID:9583949  
A;Accession: JC6552  
A;Molecule type: mRNA  
A;Residues: 1-1015 <CZE>  
A;Cross-references: GB:AF023910; GB:U63217; NID:g2642492; PIDN:AAC14193.1; PID:g2642493  
C;Genetics:  
A;Gene: top1  
C;Superfamily: eukaryotic type I DNA topoisomerase  
C;Keywords: isomerase  
F;974/Active site: Tyr #status predicted

Query Match 42.2%; Score 70; DB 2; Length 1015;  
Best Local Similarity 42.8%; Pred. No. 5.9; Matches 18; Conservative 6; Mismatches 8; Indels 10; Gaps 1;

QY 2 KKYAKKEKAAYKYE-----AKAAAEAAAKEAAY 33  
||| ||| ||| ||| ||| ||| :  
DB 293 KKEVKEDTAKOVKVEKXETPKTKPAKRKAESSESDP 334

RESULT 11  
T34081  
hypothetical protein CO2F12.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34081  
R;Miller, N.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid CO2F12.  
A;Reference number: Z21473  
A;Accession: T34081  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1130 <MIL>  
A;Cross-references: EMBL:U41545; PIDN:AAA83190.1; CESP:CO2F12.7  
C;Genetics:  
A;Gene: CESP:CO2F12.7  
A;Introns: 11/2; 118/1; 156/2; 239/2; 308/2; 398/1; 458/3; 591/1; 640/3; 756/2; 790/2; 8

Query Match 42.2%; Score 70; DB 2; Length 1130;  
Best Local Similarity 51.4%; Pred. No. 6.4; Matches 18; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY 6 KKEKAAYKAYKKEAKAKA-----AEAAAEAAAYE 34  
|:::| ::| |::| |::| |  
DB 1016 KREFAEKRTKEAEKKEKAEKEKEKAEEAKE 1050

RESULT 12  
JV0057  
tola protein - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002  
C;Accession: JV0057; B64810  
R;Levengood, S.K.; Webster, R.E.  
J. Bacteriol. 171, 6600-6609, 1989  
A;Title: Nucleotide sequences of the tola and tolB genes and localization of their prod  
A;Reference number: JV0057; MUID:90078104; PMID:2687247  
A;Accession: JV0057  
A;Molecule type: DNA  
A;Residues: 1-421 <LEV>  
A;Cross-references: GB:M28232; NID:gi48018; PIDN:AAA24683.1; PID:gi48019  
A;Experimental source: strain JM105  
A;Note: the authors translated the initiation codon GTG for residue 1 as Val  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: B64810  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-421 <BLAT>

A; Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC7383.1; PID:g1786960;  
A; Experimental source: strain K-12, substrain MG1655  
C; Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach the  
C; Genetics:

A;Gene: tola  
A;Map position: 17 min  
A;Start codon: GTG  
C;Keywords: nucleotide binding; P-loop; transmembrane protein  
F;14-34/Domain: transmembrane #status predicted <MSS>  
F;78-301/Domain: helical #status predicted <HSS>  
F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 41.9%; Score 69.5; DB 2  
Best Local Similarity 52.6%; Pred. No. 3.2;  
Matches 20; Conservative 4; Mismatches 9

RESULT 13  
F90725  
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain C; Species: Escherichia coli  
C; Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C; Accession: F90725  
R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A; fitte: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:21156231; PMID:11258796

```

Query Match      40.4%; Score 67; DB 2; Length 394;
Best Local Similarity 48.9%;
Matches 22; Conservative 5; Mismatches 8; Indels 10; Gaps 2

QY      1 AKKYAKKE--KAQKAYKK-----EAKAKAAEAAAKEAAVEA 35
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      167 AKKGAEEAAKAAAEQAQKAAAAAALKKGAEEAAAEAAAEARKK 211

```

RESULT 14  
G85576  
membrane spanning protein Tola [imported] - Escherichia coli (strain O157:H7, substrain C.) Species: Escherichia coli  
C.Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Nov-2001  
C.Accession: G85576  
R.;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A.Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A.Reference number: A85480; PMID:21074935; PMID:11206551  
A.Accession: G85576  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-394 <STO>  
A.Cross-references: GB:AE005174; NID:gr12513672; PIDN:AAGS5075.1; GSPDB:GN00145; UWGP:209  
A.Experimental source: strain O157:H7, substrain EDL933  
C.Genetics:  
A.Gene: tola

Query Match 40.4%; Score 67; DB 2; Length 394;  
Best Local Similarity 48.9%; Pred. No. 5.6;  
Matches 22; Conservative 5; Mismatches 8; Indels 10; Gaps 2;

Qy 1 AKKYAKKE--KAARKAYKK-----EAKKAAAEAAAEEAAYEA 35  
||| : ||| : ||| : ||| : ||| :  
Db 167 AKKAAEAAKAAAAEQKCAEAAAALKKCAEAAAAEAAARCKA 211

RESULT 15

G75403  
DNA topoisomerase I - Deinococcus radiodurans (strain R1)  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C/Accession: G75403  
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; D.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zal.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286 1571-1577 1999

RESULT 15  
G75403  
DNA topoisomerase I - Deinococcus radiodurans (strain R1)  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C/Accession: G75403  
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; D  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zal  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286 1571-1577 1999

```

Query Match      40.4%; Score 67; DB 2; Length 1021;
Best Local Similarity 51.4%; Pred. No. 12;
Matches 18; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 AKKYAKKEKAACKAYKKEKAKAKAAEAAAKEAAVEA 35
Db 922 ASKTAPKKAACKPAKKAAPKKAASAKTPAAKA 956

```

Search completed: March 10, 2003, 12:28:25  
Job time : 12.2637 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 7.97468 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-1  
Perfect score: 166  
Sequence: 1 AKYAKKEKAAYKKAKEAAKAAEAAEAYEA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	84	50.6	384	1	TMPE TREPH
2	76	45.8	347	1	TOLA_PSEAE
3	74	44.6	372	1	TOLA_HAEIN
4	72	43.4	210	1	H1_LYTPI
5	71.5	43.1	208	1	DBH MYCSM
6	71.5	43.1	211	1	H1B_STRPU
7	70	42.2	1130	1	YL17_CAEEL
8	69.5	41.9	421	1	TOLA_ECOLI
9	68.5	41.3	87	1	H1C6_TRYCR
10	68.5	41.3	97	1	H1C2_TRYCR
11	68.5	41.3	101	1	ASR_ENTCL
12	66.5	40.1	217	1	H1G_STRPU
13	66.5	40.1	568	1	ZRF1_HUMAN
14	66	39.8	1391	1	MST2_DROHY
15	65.5	39.5	214	1	DBH MYCTU
16	65.5	39.5	1528	1	SPAA_STRDO
17	64.5	38.9	185	1	H1D_STRPU
18	64.5	38.9	205	1	DBH_MYCBO
19	64	38.6	80	1	H161_TRYCR
20	64	38.6	90	1	H162_TRYCR
21	64	38.6	233	1	H11_GLYBA
22	64	38.6	240	1	H12_VOLCA
23	63.5	38.3	514	1	ZRF1_MOUSE
24	63.5	38.3	707	1	HS88_NEUCR
25	63	38.0	248	1	H1_PARAN
26	63	38.0	260	1	H11_VOLCA
27	62.5	37.7	232	1	H1A_CHITE
28	62.5	37.7	349	1	RS6_AEDAL
29	62	37.3	206	1	H1_ONCMY
30	62	37.3	237	1	H1E_CHITE
31	62	37.3	244	1	H1O_CHITH
32	62	37.3	898	1	IF2_VIBCH
33	61.5	37.0	233	1	H11_GLYSA

34	61.5	37.0	373	1	SECF_STRCO
35	61.5	37.0	917	1	IF2_PROVU
36	61.5	37.0	952	1	TOPI_STRCO
37	61	36.7	134	1	RS16_BRUME
38	61	36.7	202	1	H1L_MYTR
39	61	36.7	299	1	RL22_DROME
40	61	36.7	441	1	IF5_CAEEL
41	61	36.7	1093	1	SYV_NEUCR
42	60.5	36.4	184	1	RS16_BACTN
43	60.5	36.4	348	1	CYL2_HUMAN
44	60	36.1	193	1	H1O_MOUSE
45	60	36.1	193	1	H1O_RAT

## ALIGNMENTS

RESULT 1  
TMPE TREPH STANDARD; PRT; 384 AA.  
AC P29720;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Treponema membrane protein B precursor (Antigen tmpB).  
GN TMPE.  
OS Treponema phagedenis.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=162;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN=Kazan 5;  
RX MEDLINE=91372983; PubMed=1894368;  
RA Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,  
RA Slivienky L., Schullis L.M., van Embden J.D., Charon N.W.;  
RT Treponema phagedenis encodes and expresses homologs of the Treponema  
RL pallidum TmpA and TmpB proteins.;  
RL Infect. Immun. 59:3685-3693(1991).  
CC -!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR  
CC LARGE MOLECULES.  
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.  
CC -!- SIMILARITY: TO TMPE OF T.PALLIDUM.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M58563; AAA27480.1; -.  
DR PIR; B43592; B43592.  
KW Antigen; Outer membrane; Repeat; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 384  
FT DOMAIN 151 235  
FT REPEAT 151 155  
FT REPEAT 156 160  
FT REPEAT 161 165  
FT REPEAT 166 170  
FT REPEAT 171 175  
FT REPEAT 176 180  
FT REPEAT 181 185  
FT REPEAT 186 190  
FT REPEAT 191 195  
FT REPEAT 196 200  
FT REPEAT 201 205  
FT REPEAT 206 210  
FT REPEAT 211 215  
FT REPEAT 216 220  
FT REPEAT 221 225  
FT REPEAT 17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR] -  
[ED].  
1-1.  
1-2.  
1-3.  
1-4.  
1-5.  
1-6.  
1-7.  
1-8.  
1-9.  
1-10.  
1-11.  
1-12.  
1-13.  
1-14.  
1-15.

```
FT REPEAT 226 230 1-16.
FT REPEAT 231 235 1-17.
FT DOMAIN 236 288 6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-
A-A-E.
FT REPEAT 236 243 2-1.
FT REPEAT 245 252 2-2.
FT REPEAT 254 261 2-3.
FT REPEAT 263 270 2-4.
FT REPEAT 272 279 2-5.
FT REPEAT 281 288 2-6.
SQ SEQUENCE 384 AA; 42677 MW; 6E94CBC74294DE8C CRC64;

Query Match 50.6%; Score 84; DB 1; Length 384;
Best Local Similarity 64.9%; Pred. No. 0.059;
Matches 24; Conservative 5; Mismatches 6; Indels 2; Gaps 2;

QY 1 AKYAKKEKAK-KAYKKEAKAKAA-EAAAKEAAYEA 35
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 AREWAKEAKAKKAAKEEAARAAEAAEAAKAAEAA 254

RESULT 2
TOLA_PSEAE
ID TOLA_PSEAE STANDARD; PRT; 347 AA.
AC P50600;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97113525; PubMed=8955385;
RA Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT "Identification and characterization of the tolQRA genes of
Pseudomonas aeruginosa."
RL J. Bacteriol. 178:7059-7068 (1996).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Duan K., Sokol P.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.B., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen."
RL Nature 406:959-964 (2000).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
(Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U39558; AAC44660.2; -.
CC DR
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;

DR EMBL; AE004530; AAG04360.1; -.
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).
FT DOMAIN 209 216 POLY-ALA.
SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 45.8%; Score 76; DB 1; Length 347;
Best Local Similarity 56.1%; Pred. No. 0.35;
Matches 23; Conservative 5; Mismatches 7; Indels 6; Gaps 2;

QY 1 AKKYAK---KEKAACKAYKK---EAKAKAAEAAAKEAAYEA 35
|||: ||||| ||||| ||||| ||||| ||||| |||||
Db 155 AKKRAEADKKAEDAKKKAEDAKKKAEEAKKKAEEA 195

RESULT 3
TOLA_HAEIN
ID TOLA_HAEIN STANDARD; PRT; 372 AA.
AC P44678; P94810;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE TOLA protein.
GN TOLA OR HI0383.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Spriggs T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd."
RL Science 269:496-512 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1479;
RX MEDLINE=97080550; PubMed=8921895;
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae tolQ,
tolR, tolA and tolB genes."
RL Gene 178:75-81 (1996).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
(Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U32722; AAC22041.1; -.
CC DR
CC EMBL; U32470; AAC44596.1; -.
CC DR
CC HSPF; P19934; ITOL.
CC TIGR; HI0383; -.
CC KW
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
```

Query Match 43.4%; Score 72; DB 1; Length 210;  
Best Local Similarity 59.0%; Pred. No. 0.57;  
Matches 23; Conservative 3; Mismatches 9; Indels 4; Gaps 2;

QY 1 AKKYAKK--EKAACK-AVKKEAKAKAEAAAKEAAYEA 35  
DB 157 AKKAARKPAKGAARPAKAPKPKAAKPAKAAKPAKAAKPAK 195

RESULT 5  
DBH\_MYCSM STANDARD; PRT; 208 AA.  
AC QZHC5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
GN HUP OR HLP.  
OS Mycobacterium smegmatis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1772;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 700084 / mc(2)155;  
RX MEDLINE=9110209; PubMed=9894918;  
RA Lee B.H., Murugasu-Oei B., Dick T.;  
RT "Upregulation of a histone-like protein in dormant Mycobacterium smegmatis";  
RL Mol. Gen. Genet. 260:475-479(1998).  
CC -I- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
CC -I- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF068138; AAD13809.1; --  
DR HSSP; P02346; 1HTU.  
DR InterPro; IPR000119; Bac\_DNABind.  
DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
DR ProDom; PD000945; Bac\_DNABind; 1.  
DR SMART; SM00411; BHL; 1.  
DR PROSITE; PS00045; HISTONE LIKE; 1.  
KW DNA-binding; DNA condensation; Repeat.  
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
SQ SEQUENCE 208 AA; 21230 MW; CA5F577F61F7EF09 CRC64;

Query Match 43.1%; Score 71.5; DB 1; Length 208;  
Best Local Similarity 58.3%; Pred. No. 0.64;  
Matches 21; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 AKKYAKKEKAAYKKEA-KAKAAEAAAKEAAYEA 35  
DB 111 AKKAARKPAKGAARPAKAPKPKAAKPAKAAKPAKAAKPAK 146

RESULT 6  
H1B\_STRPU STANDARD; PRT; 211 AA.  
AC P15869;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Histone H1-beta, late embryonic.

```
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinozoa; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88246461; PubMed=2837660;
RA Lai Z.-C., Childs G.;
RT "Characterization of the structure and transcriptional patterns of
RT the gene encoding the late histone subtype H1-beta of the sea urchin
RT Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844(1988).
CC -I- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20314; AAA30052.1; -.
CC PIR; A28100; A28100.
CC HSP; P02259; IHST.
CC InterPro; IPR001386; Histone H1/H5.
CC InterPro; IPR003216; Linkerhist N.
CC Pfam; PF00538; linker histone; 1.
CC ProDom; PD000373; Linkerhist_N; 1.
CC SMART; SM00526; H1.5; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
KW CHROMOSOMAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; MULTIGENE FAMILY.
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334BBE7A CRC64;

Query Match 43.1%; Score 71.5; DB 1; Length 211;
Best Local Similarity 66.7%; Pred. No. 0.65;
Matches 20; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 1 AKYAKKEKAAKAYKKEAKAKAAKAAK 30
DQ 183 AKKAAK-PAKKAACKVAKKPAKKAACK 211

RESULT 7
YL17_CABEL YL17_CABEL STANDARD; PRT; 1130 AA.
AC Q11102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.
GN C02F12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Miller N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: WEAK, TO MYOSINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```

```
CC EMBL; U41545; AAK39135.1; -.
DR WormPep; C02F12.7; C02901.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 121 779 COILED COIL (POTENTIAL).
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
SQ SEQUENCE 1130 AA; 131485 MW; B0PD2EFE3D99FB09 CRC64;

Query Match 42.2%; Score 70; DB 1; Length 1130;
Best Local Similarity 51.4%; Pred. No. 3.7;
Matches 18; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY 6 KKEAKAKAYKKEAKAKA-----AEAAKAAAYE 34
DQ 1016 KKEASEKRTKAEKAEKAEKAEKAEKAEKAE 1050

RESULT 8
TOLA_ECOLI TOLA_ECOLI STANDARD; PRT; 421 AA.
AC P19934;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR CIM OR EXCC OR LKY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JM105;
RX MEDLINE=90078104; PubMed=2687247;
RA Levensgood S.K., Webster R.E.;
RT "Nucleotide sequences of the tola and tolB genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP DOMAINS.
RX MEDLINE=91296736; PubMed=2068069;
RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;
RT "TolA: a membrane protein involved in colicin uptake contains an
RT extended helical region.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN [5]
RP INTERACTION WITH PORINS.
RX MEDLINE=97133271; PubMed=8978668;
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
```

RA Llobes R.;  
 RT "TOLA central domain interacts with Escherichia coli porins.";  
 RL ENBO J. 15:6408-6415 (1996).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.  
 RX MEDLINE=99332679; PubMed=10404600;  
 RA Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;  
 RT "Filamentous phage infection: crystal structure of gsp in complex  
 with its coreceptor, the C-terminal domain of TolA.";  
 RL Structure 7:711-722 (1999).  
 CC -1- FUNCTION: INVOLVED IN THE TOMB-INDEPENDENT UPTAKE OF GROUP A  
 COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE  
 COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
 BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION  
 OF BACTERIOPHAGE DNA.  
 CC -1- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE  
 AND LAMB.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M28232; AAA24683.1; --  
 DR EMBL; AE000177; AAC73833.1; --  
 DR EMBL; D90713; BAA35405.1; --  
 DR PIR; J00057; J00057.  
 DR PDB; 1TOL; 20-MAY-99.  
 DR EcoGene; EG11007; tola.  
 KW Transport; Protein transport; Bacteriocin transport; Transmembrane;  
 KW Repeat; Inner membrane; 3D-structure; Complete proteome.  
 FT DOMAIN 1 13  
 FT TRANSMEM 14 34  
 FT DOMAIN 35 421  
 FT DOMAIN 48 310  
 FT DOMAIN 311 421  
 FT DOMAIN 224 278  
 FT A(2,4).  
 FT  
 SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;  
 Query Match 41.9%; Score 69.5; DB 1; Length 421;  
 Best Local Similarity 52.6%; Pred. No. 1.8;  
 Matches 20; Conservative 4; Mismatches 9; Indels 5; Gaps 1;  
 QY 3 KYAKKEAKAKA-----YKKEAKAKAFAAAKAAEAAAYEA 35  
 Db 130 KQQAEEAAKAAADAKAKAEADAKAEAEAAKAAADA 167  
 RESULT 9  
 HIC2\_TRYCR STANDARD; PRT; 87 AA.  
 ID HIC6\_TRYCR  
 AC P40269; P40271;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Histone H1.C6/H1.C9.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tulahuén 2;  
 RX MEDLINE=95059220; PubMed=7969272;  
 RA Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,  
 RA Galanti N., Pettersson U.;  
 RT "A gene family encoding heterogeneous histone H1 proteins in  
 Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 65:317-330 (1994).  
 Query Match 41.9%; Score 69.5; DB 1; Length 421;  
 Best Local Similarity 52.6%; Pred. No. 1.8;  
 Matches 20; Conservative 4; Mismatches 9; Indels 5; Gaps 1;  
 QY 3 KYAKKEAKAKA-----YKKEAKAKAFAAAKAAEAAAYEA 35  
 Db 130 KQQAEEAAKAAADAKAKAEADAKAEAEAAKAAADA 167  
 RESULT 9  
 HIC6\_TRYCR STANDARD; PRT; 87 AA.  
 ID HIC6\_TRYCR  
 AC P40269; P40271;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Histone H1.C6/H1.C9.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tulahuén 2;  
 RX MEDLINE=95059220; PubMed=7969272;  
 RA Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,  
 RA Galanti N., Pettersson U.;  
 RT "A gene family encoding heterogeneous histone H1 proteins in  
 Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 65:317-330 (1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L27120; AAA66479.1; --  
 DR EMBL; L27117; AAA66481.1; --  
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
 SQ SEQUENCE 87 AA; 9005 MW; 843FF90F5F73922D CRC64;  
 Query Match 41.3%; Score 68.5; DB 1; Length 87;  
 Best Local Similarity 52.6%; Pred. No. 0.62;  
 Matches 20; Conservative 3; Mismatches 12; Indels 3; Gaps 1;  
 QY 1 AKKYA---KKEKAAYKAYKKEAKAKAAEAAKAAAYEA 35  
 Db 47 AKKRAAPKKPAAAKAVTKSAKKHAAKAPKAVKKA 84  
 RESULT 10  
 HIC2\_TRYCR STANDARD; PRT; 97 AA.  
 ID HIC2\_TRYCR  
 AC P40268;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Histone H1.C2.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tulahuén 2;  
 RX MEDLINE=95059220; PubMed=7969272;  
 RA Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,  
 RA Galanti N., Pettersson U.;  
 RT "A gene family encoding heterogeneous histone H1 proteins in  
 Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 65:317-330 (1994).  
 Query Match 41.3%; Score 68.5; DB 1; Length 97;  
 Best Local Similarity 52.6%; Pred. No. 0.68;  
 Matches 20; Conservative 3; Mismatches 12; Indels 3; Gaps 1;  
 QY 1 AKKYA---KKEKAAYKAYKKEAKAKAAEAAKAAAYEA 35  
 Db 57 AKKRAAPKKPAAAKAVTKSAKKHAAKAPKAVKKA 94  
 RESULT 11  
 ASR\_ENTCL STANDARD; PRT; 101 AA.  
 ID ASR\_ENTCL  
 AC 093MH6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Acid shock protein precursor.





Search completed: March 10, 2003, 12:17:07  
Job time : 8.97468 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 18.5338 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-1

Perfect score: 166

Sequence: 1 AKYAKKEKAAYKAYKKEAKAAEAAYEA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPRTEMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81	48.8	372	2 Q9WXX1	Q9wxx1 pseudomonas
2	80.5	48.5	461	2 O68124	O68124 rhodobacter
3	79.5	47.9	198	5 Q8WQ44	Q8wq44 leishmania
4	79.5	47.9	111	5 Q8T9R3	Q8t9r3 leishmania
5	77	46.4	275	5 O01395	O01395 drosophila
6	77	46.4	395	2 Q937K4	Q937k4 erwinia chr
7	75.5	45.5	101	5 Q9BMV8	Q9bmv8 leishmania
8	75	45.2	244	16 Q9AJX2	Q9ajx2 streptomyce
9	70.5	42.5	445	3 Q9PQJ8	Q9pqj8 neurospora
10	70	42.2	277	16 Q9XAQ3	Q9xaq3 streptomyce
11	70	42.2	1015	5 Q94705	Q94705 physarum po
12	69.5	41.9	78	5 Q95UN5	Q95un5 trypanosoma
13	69	41.6	741	5 Q9X207	Q9x207 drosophila
14	68.5	41.3	83	5 Q95UN6	Q95un6 trypanosoma
15	68.5	41.3	112	5 Q9XYI5	Q9xyi5 leishmania
16	68.5	41.3	629	2 Q9KJ98	Q9kj98 escherichia

17	68	41.0	241	5 Q17321	Q17321 chironomus
18	68	41.0	467	2 Q9AGV7	Q9agv7 corynebacte
19	68	41.0	1092	16 Q98FB4	Q98fb4 rhizobium l
20	67.5	40.7	81	5 Q9NFP6	Q9nfp6 trypanosoma
21	67	40.4	394	16 Q8X965	Q8x965 escherichia
22	67	40.4	515	5 Q9VBL3	Q9vbl3 drosophila
23	67	40.4	1021	16 Q9RULO	Q9rul0 deinococcus
24	66.5	40.1	71	5 Q9NFI8	Q9nfi8 trypanosoma
25	66.5	40.1	76	5 Q9NFI8	Q9nfi8 trypanosoma
26	66.5	40.1	76	5 Q9NFI8	Q9nfi8 trypanosoma
27	66.5	40.1	81	5 Q9NFI8	Q9nfi8 trypanosoma
28	66.5	40.1	81	5 Q9NFI8	Q9nfi8 trypanosoma
29	66.5	40.1	582	4 O60415	O60415 homo sapien
30	66.5	40.1	635	4 O60414	O60414 homo sapien
31	66.5	40.1	969	5 Q9NDI9	Q9ndi9 plasmodium
32	66.5	40.1	1052	5 Q26938	Q26938 trypanosoma
33	66	39.8	224	13 Q90ZD7	Q90zd7 bufo bufo g
34	66	39.8	232	10 Q39576	Q39576 chlamydomon
35	66	39.8	300	10 Q8VWK4	Q8vwk4 arabidopsis
36	66	39.8	318	10 Q9LPL9	Q9lpl9 arabidopsis
37	65.5	39.5	67	16 Q8XWI3	Q8xwi3 ralstonia s
38	65.5	39.5	481	5 Q9VWK3	Q9vwk3 drosophila
39	65.5	39.5	612	10 Q9XHE2	Q9xhe2 chlamydomon
40	65	39.2	101	2 Q9X342	Q9x342 bacillus an
41	65	39.2	322	5 Q9VFI1	Q9vfi1 drosophila
42	65	39.2	497	2 Q45302	Q45302 corynebacte
43	65	39.2	498	2 Q9ANT7	Q9ant7 brevivbacte
44	65	39.2	901	5 Q8SUW1	Q8suw1 encephalito
45	64.5	38.9	191	5 O46142	O46142 mytilus edu

#### ALIGNMENTS

#### RESULT 1

ID	Q9WXX1	PRELIMINARY;	PRT;	372 AA.
AC	Q9WXX1;			
DT	01-NOV-1999 (TREMREL. 12, Created)			
DT	01-NOV-1999 (TREMREL. 12, Last sequence update)			
DT	01-MAR-2002 (TREMREL. 20, Last annotation update)			
DE	Tola protein.			
GN	TOLA.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=303;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96198174; PubMed=8626299;			
RA	Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;			
RT	"The Pseudomonas putida peptidoglycan-associated outer membrane			
RT	lipoprotein (PAL) is involved in maintenance of the integrity of the			
RT	cell envelope."			
RL	J. Bacteriol. 178:1699-1706(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Ramos-Gonzalez I.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RA	Rodriguez-Herva J.J.;			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT-2;			
RX	MEDLINE=96422022; PubMed=8824639;			
RA	Rodriguez-Herva J.J., Ramos J.;			
RT	"Characterization of an OprL null mutant of Pseudomonas putida."			
RL	J. Bacteriol. 178:5836-5840(1996).			

```

DR EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PRO0624; HISTONEHS.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match
Best Local Similarity 56.1%; Pred. No. 0.21;
Matches 23; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

Qy 1 AKKYA-----KKEAKKAYKKEAKAKAAEAAAEAAAYEA 35
Db 160 AKKADEAKKAAEEAAKAAAEAAEAKKAAAEADAKKAAEEA 200

RESULT 2
O68124 PRELIMINARY; PRT; 461 AA.
AC O68124;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Hypothetical 49.5 kDa protein.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=97404404; PubMed=9256491;
RA Vilek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter
RT capsulatus SB1003."
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
DR EMBL; AF010496; AAC16214.1; -.
KW Hypothetical protein.
SQ SEQUENCE 461 AA; 49516 MW; DD8DA03418BC0368 CRC64;

Query Match
Best Local Similarity 48.5%; Score 80.5; DB 2; Length 461;
Matches 21; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

Qy 1 AKKYAKKEAKKAYKKEAKAKAAEAAAEAAAYEA 35
Db 30 AKEWEKEAAAKDAEAAKAKAR-EEAAKDAAKA 63

RESULT 3
Q8WQ44 PRELIMINARY; PRT; 98 AA.
AC Q8WQ44;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE 18 kDa nuclear protein.
GN LNP18.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RA Tzortzis N., Papageorgiou F.T., Tzinia A.K., Soteriadou K.P.;
RT "Identification and characterization of a novel Leishmania gene
RT encoding for a putative histone H1-like transcription factor."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237814; CAD21431.1; -.
KW Nuclear protein.
SQ SEQUENCE 98 AA; 9999 MW; 0A4AB93089D6C261 CRC64;

Query Match
Best Local Similarity 47.9%; Score 79.5; DB 5; Length 98;
Matches 23; Conservative 3; Mismatches 5; Indels 3; Gaps 2;

DR EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PRO0624; HISTONEHS.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match
Best Local Similarity 56.1%; Pred. No. 0.21;
Matches 23; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

Qy 1 AKKYA-----KKEAKKAYKKEAKAKAAEAAAEAAAYEA 35
Db 160 AKKADEAKKAAEEAAKAAAEAAEAKKAAAEADAKKAAEEA 200

RESULT 2
O68124 PRELIMINARY; PRT; 461 AA.
AC O68124;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE Hypothetical 49.5 kDa protein.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=97404404; PubMed=9256491;
RA Vilek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter
RT capsulatus SB1003."
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
DR EMBL; AF010496; AAC16214.1; -.
KW Hypothetical protein.
SQ SEQUENCE 461 AA; 49516 MW; DD8DA03418BC0368 CRC64;

Query Match
Best Local Similarity 48.5%; Score 80.5; DB 2; Length 461;
Matches 21; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

Qy 1 AKKYAKKEAKKAYKKEAKAKAAEAAAEAAAYEA 35
Db 30 AKEWEKEAAAKDAEAAKAKAR-EEAAKDAAKA 63

RESULT 3
Q8WQ44 PRELIMINARY; PRT; 98 AA.
AC Q8WQ44;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE 18 kDa nuclear protein.
GN LNP18.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RA Tzortzis N., Papageorgiou F.T., Tzinia A.K., Soteriadou K.P.;
RT "Identification and characterization of a novel Leishmania gene
RT encoding for a putative histone H1-like transcription factor."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237814; CAD21431.1; -.
KW Nuclear protein.
SQ SEQUENCE 98 AA; 9999 MW; 0A4AB93089D6C261 CRC64;

Query Match
Best Local Similarity 47.9%; Score 79.5; DB 5; Length 98;
Matches 23; Conservative 3; Mismatches 5; Indels 3; Gaps 2;

Qy 1 AKKYAKK--EKAACKAYKKEAKAKAAEAAAEAA 32
Db 63 AKKVAKKPAKKAACKKPAKPAK-KAACKAAKKA 95

RESULT 4
Q8T9R3 PRELIMINARY; PRT; 111 AA.
AC Q8T9R3;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE 18 kDa nuclear protein.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RA Papageorgiou F., Soteriadou K.;
RT "Identification of a Leishmania infantum gene encoding for an histone
RT H1-like nuclear protein."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF469106; AAL76335.1; -.
KW Nuclear protein.
SQ SEQUENCE 111 AA; 11162 MW; 16168F3B54960E83 CRC64;

Query Match
Best Local Similarity 47.9%; Score 79.5; DB 5; Length 111;
Matches 23; Conservative 3; Mismatches 5; Indels 3; Gaps 2;

Qy 1 AKKYAKK--EKAACKAYKKEAKAKAAEAAAEAA 32
Db 76 AKKVAKKPAKKAACKKPAKPAK-KAACKAAKKA 108

RESULT 5
O01395 PRELIMINARY; PRT; 275 AA.
AC O01395;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE Axoneme-associated protein MST101(3).
GN MST101(3) OR DHMST101..
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A.
RA Neesen J., Heinlein U.A.O., Buenemann H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -.
DR FlyBase; FBgn020732; Dhyd\mst101(3).
KW Sperm; Repeat; Multigene family.
FT DOMAIN 64 255
FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-
FT [AE]-X
SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2ADF732C CRC64;

Query Match
Best Local Similarity 46.4%; Score 77; DB 5; Length 275;
Matches 20; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 AKKYAKKAAKAYKKEAKAKAAEAAAEAAAYEA 35
Db 196 AKKAAEEAAKKAAEEVAACKKADAEAAKKAEEA 230

```

RESULT 10  
Q9XAQ3



Search completed: March 10, 2003, 12:25:52  
Job time : 20.5338 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 18.2384 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-1  
Perfect score: 166  
Sequence: 1 AKYAKKEKAANKYKKEAKAKAAEAARAAEAAYEA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166	100.0	35	21	Copolymer molecule
2	108	65.1	45	21	Copolymer molecule
3	108	65.1	86	21	Copolymer molecule
4	105.5	63.6	66	21	Copolymer molecule
5	105	63.3	109	21	Copolymer molecule
6	104.5	63.0	56	21	Copolymer molecule
7	104.5	63.0	77	21	Copolymer molecule
8	86.5	52.1	427	22	C albicans apoptos
9	76	45.8	154	11	Recombinant copoly
10	73.5	44.3	40	21	Peptide modulating

11	71	42.8	106	11	Recombinant copoly
12	69.5	41.9	146	22	Human protein sequ
13	69.5	41.9	165	22	C glutamicum prote
14	69	41.6	741	22	Drosophila melanog
15	67.5	40.7	32	16	Polycationic polyp
16	67.5	40.7	32	18	Protamine-like pep
17	67.5	40.7	32	21	Peptide modulating
18	67.5	40.7	33	16	Polycationic polyp
19	67.5	40.7	33	18	Protamine-like pep
20	67	40.4	515	22	Drosophila melanog
21	66.5	40.1	582	23	Human endometrial
22	66	39.8	242	21	Arabidopsis thalia
23	66	39.8	242	21	Arabidopsis thalia
24	66	39.8	245	21	Arabidopsis thalia
25	66	39.8	245	21	Arabidopsis thalia
26	66	39.8	300	21	Arabidopsis thalia
27	66	39.8	300	21	Arabidopsis thalia
28	65.5	39.5	214	20	M. tuberculosis hi
29	65.5	39.5	214	21	M. tuberculosis hi
30	65.5	39.5	481	22	Drosophila melanog
31	65	39.2	223	20	Amino acid sequenc
32	65	39.2	223	22	Drosophila melanog
33	65	39.2	223	22	C glutamicum prote
34	64.5	38.9	203	18	Streptococcus pneu
35	64.5	38.9	205	21	Mycobacterium bovi
36	64.5	38.9	769	22	Enterococcus faeca
37	64.5	38.9	782	22	Enterococcus faeca
38	63.5	38.3	621	20	Breast cancer asso
39	63	38.0	607	22	Drosophila melanog
40	63	38.0	2451	22	Drosophila melanog
41	62.5	37.7	29	16	Polycationic polyp
42	62.5	37.7	29	16	Polycationic polyp
43	62.5	37.7	29	18	Protamine-like pep
44	62.5	37.7	29	18	Protamine-like pep
45	62.5	37.7	262	17	T. cruzi L19E homo

ALIGNMENTS

RESULT 1

AA082571

ID AA082571 standard; peptide; 35 AA.

XX AA082571;

XX AC

XX DT 28-JUL-2000 (first entry)

XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:1.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease; glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective; osteopathic; immunosuppressive; antithyroid; antiinflammatory; antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological; antianemic; immunosuppressive; demyelinating disease; osteoarthritis; inflammatory condition; multiple sclerosis; rheumatoid arthritis; Crohn's disease; chronic immune thrombocytopaenia purpura; colitis; diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis; Hashimoto's disease; idiopathic myxoedema; myasthenia gravis; pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX OS

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX YEDA ) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.





[illegible]



RESULT 7  
 AAY82575  
 ID AAY82575 standard; peptide; 77 AA.  
 XX  
 AC AAY82575;  
 XX  
 DT 28-JUL-2000 (first entry)  
 XX  
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.  
 XX  
 KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 XX  
 OS Unidentified.  
 XX  
 XX WO200018794-A1.  
 XX  
 PD 06-APR-2000.  
 XX  
 XX 24-SEP-1999; 99WO-US22402.  
 XX  
 XX 25-SEP-1998; 98US-0101693.  
 XX  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 XX (TEVA-) TEVA PHARM USA INC.  
 PA  
 PA  
 XX Gad A, Lis D;  
 PI  
 XX WPI; 2000-317499/27.  
 DR  
 XX  
 XX Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glatiramer acetate and for treatment and prevention of immune diseases  
 XX  
 XX Claim 10; Page 14; 72pp; English.  
 PS  
 XX  
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions,  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights, and physical properties which are analogous to  
 CC glatiramer acetate molecules, which makes them ideal for use as  
 CC molecular weight markers.  
 XX  
 SQ Sequence 77 AA;

Query Match 63.0%; Score 104.5; DB 21; Length 77;  
 Best Local Similarity 65.8%; Pred. No. 2.9e-05;  
 Matches 25; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY 1 AKKYAKKEKAAYKAY---KKEAKAKAAEAEEAAKEAAVEA 35  
 ||||| ||||| : |||||  
 Db 40 AKKYAKAAKEKKEKAYAAAEAKYKAAEAKEAAKEAAVEA 77  
 ||||| ||||| : |||||  
 RESULT 8  
 AAG70868  
 ID AAG70868 standard; Protein; 427 AA.  
 XX  
 AC AAG70868;  
 XX  
 DT 27-JUL-2001 (first entry)  
 XX  
 DE C albicans apoptosis associated protein #48.  
 XX  
 KW Yeast; fungus; apoptosis; infection; proliferative disease;  
 KW vaccine; autoimmune disease; ischaemia; neurodegeneration.  
 XX  
 OS Candida albicans.  
 XX  
 PN WO200102550-A2.  
 XX  
 XX 11-JAN-2001.  
 PD  
 XX 03-JUL-2000; 2000WO-BE00077.  
 PF  
 XX 01-JUL-1999; 99EP-0870141.  
 PR  
 XX (JANC ) JANSSEN PHARM NV.  
 PA  
 XX Contreras RH, De Backer MD, Luyten WHML, Malcorps IXL;  
 PI Nelissen BJW, Reekmans RJ;  
 PI  
 XX WPI; 2001-367042/38.  
 DR N-PSDB; AAH29904.  
 DR  
 XX Yeast and fungal nucleic acids encoding proteins involved in a pathway  
 PT leading to programmed cell death, useful for treating proliferative  
 PT disorders, yeast and fungal infections, or for preventing apoptosis in  
 PT certain diseases -  
 XX  
 PS Claim 24; Fig 2; 218pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of apoptosis associated proteins from the yeast *Saccharomyces*  
 CC *cerevisiae* and the fungus *Candida albicans*. These can be used to identify  
 CC treatments for fungal and yeast infections, for proliferative diseases  
 CC and for apoptosis related diseases such as autoimmune diseases, ischaemia  
 CC and neurodegeneration. The present sequence is one of the *C. albicans*  
 CC proteins of the invention.  
 XX  
 SQ Sequence 427 AA;  
 Query Match 52.1%; Score 86.5; DB 22; Length 427;  
 Best Local Similarity 65.7%; Pred. No. 0.02;  
 Matches 23; Conservative 3; Mismatches 8; Indels 1; Gaps 1;  
 QY 2 KKYAKKEKAAYKKEAKAKAA-EEAAKEAAVEA 35  
 ||||| ||||| : |||||  
 Db 283 KEKAKKEKAAYKKEKSGSRKAAEEAAKAAEEA 317  
 ||||| ||||| : |||||  
 RESULT 9  
 AAR06445  
 ID AAR06445 standard; protein; 154 AA.  
 XX  
 AC AAR06445;  
 XX  
 DT 03-JAN-1991 (first entry)  
 XX  
 DE Recombinant copolymer 1-77, myelin basic protein analogue.  
 XX

KW Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
 KW immunological activity; autoimmune encephalomyelitis;  
 KW multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX PD 22-AUG-1990.

XX PF 16-FEB-1990; 90EP-0301700.

XX PR 07-FEB-1990; 90US-0473845.

XX PR 17-FEB-1989; 89US-0312541.

XX PA (REPL-) REPLIGEN CORP.

XX PI Cook KS;

XX DR WPI; 1990-255848/34.

XX DR N-PSDB; AAQ05664.

XX PT Producing genes encoding random polymers of aminoacid(s) - for  
 PT producing recombinant polypeptide(s) with biological and/or  
 PT immunological activity

XX PS Disclosure; Fig 11; 25pp; English.

XX CC To improve the expression of rCOP-1 polypeptides in E. coli, genes  
 CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN  
 CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
 CC express Protein A. The resulting plasmids encode fusion proteins  
 CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
 CC A methionine residue occurs between the Protein A and rCOP-1  
 CC sequences, originating from the 5' linker sequence, in order that  
 CC the COP-1 polypeptide may be cleaved from the fusion protein.  
 CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
 CC segments: YKK, EAE, KAK, and AAA. The N-terminal alanine residue  
 CC is left behind following CNBr cleavage of the fusion protein.  
 CC The product prevents or arrests experimental autoimmune  
 CC encephalomyelitis. They are used to prevent, arrest or control a  
 CC demyelinating disorder, e.g. multiple sclerosis. They may also  
 CC be used as additives to hair care products to confer beneficial  
 CC effects on damaged hair or as supplements for diets deficient in  
 CC certain amino acids.  
 CC See also AAQ05665.

XX SQ Sequence 154 AA;

Query Match . 45.8%; Score 76; DB 11; Length 154;  
 Best Local Similarity 54.1%; Pred. No. 0.12;  
 Matches 20; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

OY 2 KKYAKKEAKKA---YKKEAKAKAEAAKEAAE 34

Db 24 KKAKEAEKAKYKYYKKEAEAKAKAAKAAAYK 60

RESULT 10

AAB08170

XX ID AAB08170 standard; peptide; 40 AA.

XX AC AAB08170;

XX DT 04-DEC-2000 (first entry)

XX DE Peptide modulating activity of heparin, and other glycans.

XX KW Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;  
 KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;  
 KW cartilage differentiation; wound healing.

XX OS Synthetic.

XX WO200045831-A1.

XX PD 10-AUG-2000.

XX PF 02-FEB-2000; 2000WO-US02853.

XX PR 02-FEB-1999; 99US-0118276.

XX PA (UYJE-) UNIV JEFFERSON THOMAS.

XX PI San Antonio JD, Verrecchio A, Schick BP;

XX DR WPI; 2000-543446/49.

XX PT Novel synthetic peptides with high affinity for glycoaminoglycans and  
 PT proteoglycans, useful for modulating heparin, promoting cell  
 PT attachment, modulating tumour metastasis and modulating wound healing -  
 XX PS Disclosure; Page 30; 76pp; English.

XX CC The present sequence represents a synthetic peptide which has a high  
 CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful  
 CC in methods for modulating heparin or other glycoaminoglycans with  
 CC anticoagulant activity, promoting cell attachment or adhesion to  
 CC natural or synthetic surfaces (especially vein grafts), modulating  
 CC tumour cell metastasis, modulating cartilage differentiation, targeting  
 CC drugs to epithelial cell surfaces (or to other cells expressing  
 CC proteoglycans), modulating enzymes that act on glycoaminoglycan  
 CC substrates, affinity purification of bioactive sequences of a  
 CC glycoaminoglycan, modifying endothelial cell pro-coagulant or  
 CC anti-coagulant functions mediated through glycoaminoglycans, and  
 CC modulating wound healing. The peptide may also be used for blocking  
 CC tissue uptake of heparin or other glycoaminoglycans in a mammal to  
 CC increase heparin half-life in circulation.

XX SQ Sequence 40 AA;

Query Match 44.3%; Score 73.5; DB 21; Length 40;  
 Best Local Similarity 61.1%; Pred. No. 0.061;  
 Matches 22; Conservative 4; Mismatches 5; Indels 5; Gaps 2;

OY 2 KKYAK--KKAARKKAYKKEAKA---KAAEAAAKKEAA 32

Db 3 KKAARKKAAKKAARKKAAKKAARKKAAKKAARKKAA 38

RESULT 11

AAR06446

XX ID AAR06446 standard; protein; 106 AA.

XX AC AAR06446;

XX DT 03-JAN-1991 (first entry)

XX DE Recombinant copolymer 1-19, myelin basic protein analogue.

XX KW Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;  
 KW immunological activity; autoimmune encephalomyelitis;  
 KW multiple sclerosis;

XX OS Synthetic.

XX PN EP383620-A.

XX PD 22-AUG-1990.

XX PF 16-FEB-1990; 90EP-0301700.

XX PR 07-FEB-1990; 90US-0473845.

XX PR 17-FEB-1989; 89US-0312541.

XX PA (REPL-) REPLIGEN CORP.





xx Novel peptides are claimed which contain 20-40 amino acids and  
cc which have a total cationic charge of less than +21 (pref. +16 to +18;  
cc esp. +18) as determined by the number of positively charged amino acids  
cc in the sequence. Preferably the positive charges are grouped in clusters  
cc separated by neutral amino acids. Especially the peptides contain 28-32  
cc amino acids having blocks of 2-4 positively charged amino acids  
cc separated by blocks of 2-6 neutral acids. Alternatively the positive  
cc charge may be distributed evenly or randomly along the peptide sequence.  
cc In particular the peptides are analogues of n-protamine (total cationic  
cc charge = +21) in which selected arginine residues have been replaced  
cc with uncharged amino acids and other arginine residues have been  
cc replaced by other positively charged amino acid residues, preferably  
cc lysine. The peptides reverse the effect of low mol. wt. heparin (LMWH)  
cc anticoagulation and hence can be used medically to prevent bleeding  
cc after the conclusion of clinical procedures using heparin therapy. They  
cc are less toxic than n-protamine since the reduced positive charge gives  
cc an improved efficiency to toxicity ratio; and they may be more effective  
cc than n-protamine in their anti-LMWH action.  
cc The present sequence (total cationic charge = +18) is a specific example  
cc of the new polypeptides.

xx SQ Sequence 32 AA;

Query Match 40.7%; Score 67.5; DB 16; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.24;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKKYAKK-EKAACKAYKKEAKAKAEAAAKE 30  
||| ||| : ||||| | ||| ||| :  
Db 2 AKKAACKAKKAAKKAACKAKKAACKAKK 32

Search completed: March 10, 2003, 12:21:27  
Job time : 19.2384 secs

**THIS PAGE BLANK (USPTO)**





```

; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-2

Query Match          65.1%; Score 108; DB 10; Length 45;
Best Local Similarity 64.4%; Pred. No. 1.6e-06;
Matches 29; Conservative 1; Mismatches 5; Indels 10; Gaps 2;

QY 1 AKYAKKEKA--AKYAK-----KEAKAKAEAAAEAAAYEA 35
      ||||| || ||||| || ||||| ||||| ||||| |||||
Db 1 AKYAKKAEKAKAYAKAEAKKAYKAEAKKAEKAEKAEAAAYEA 45
      ||||| || ||||| || ||||| ||||| ||||| |||||

RESULT 3
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match          65.1%; Score 108; DB 10; Length 86;
Best Local Similarity 57.4%; Pred. No. 3.1e-06;
Matches 27; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 AKYAKKEKA--AKYAK-----KKEAKAKAEAAAEAAAYEA 35
      ||||| || ||||| || ||||| ||||| ||||| |||||
Db 40 AKYAKKAEKAEKAYKAEAKKAYKAEAKKAYKAEAKKAEAAAYEA 86
      ||||| || ||||| || ||||| ||||| ||||| |||||

RESULT 4
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US

```

```

; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match          63.6%; Score 105.5; DB 10; Length 66;
Best Local Similarity 43.9%; Pred. No. 4.5e-06;
Matches 29; Conservative 2; Mismatches 4; Indels 31; Gaps 1;

QY 1 AKYAKKEKA--AKYAK-----KAKAEAAAEAAAK 29
      ||||| || ||||| || ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYAKAEAKKAEAKKAEAKKAEAKKAEKAEAAAEAAKAEAAKAAK 60
      ||||| || ||||| || ||||| ||||| ||||| |||||

QY 30 EAAVEA 35
      ||||| |||||
Db 61 EAAVEA 66
      ||||| |||||

RESULT 5
US-09-816-989A-7
; Sequence 7, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match          63.3%; Score 105; DB 10; Length 109;
Best Local Similarity 62.2%; Pred. No. 8.8e-06;
Matches 28; Conservative 2; Mismatches 5; Indels 10; Gaps 2;

QY 1 AKYAKKEKA--AKYAK-----KKEA-----KAKAEAAAEAAAYEA 35
      ||||| || ||||| || ||||| ||||| ||||| |||||
Db 65 AKYAKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEAAAYEA 109
      ||||| || ||||| || ||||| ||||| ||||| |||||

RESULT 6
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE

```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 372
; TYPE: PRT
; ORGANISM: H. influenzae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: outer membrane integrity protein (tolA)
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|1573353
US-09-820-843A-8

Query Match      44.6%; Score 74; DB 9; Length 372;
Best Local Similarity 58.1%; Pred.No. 0.12;
Matches 18; Conservative 5; Mismatches 8; Indels 0; Gaps 1;

Qy    5 AKKEKAACKAYKKEAKAKAAEAEEAAKEAAEYEA 35
      ||:|||||||:||:|||||||:|:|:|
Db   158 AKRLAAAKQAEEAEKAKAAETIAAQKAKQEAA 188

RESULT 9
US-09-738-626-5751
; Sequence 5751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5751
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

Query Match      41.9%; Score 69.5; DB 9; Length 165;
Best Local Similarity 50.0%; Pred.No. 0.17;
Matches 19; Conservative 6; Mismatches 8; Indels 5; Gaps 1;

Qy    2 KKYAKEKAACKAYKKEAKAKAA-----EAAAKEAAEYEA 34
      ||:|||||||:|:|:|
Db   126 KKAREDKEAKEAAEKAAAEKAAAEESEAPAEBAEAAE 163

RESULT 10
US-09-919-497-100
; Sequence 100, Application US/09919497
; Patent No. US2002010662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
```



;  
;  
PRIOR APPLICATION DATA:

**THIS PAGE BLANK (USPTO)**



;  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-152-488-13

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKKAACKAKKAACKAKKAACKAKKAACK 32

## RESULT 2

US-08-303-025-15  
; Sequence 15, Application US/08303025  
; Patent No. 5614494  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 150 West Jefferson, Suite 2500  
; CITY: Detroit  
; STATE: Michigan  
; COUNTRY: United States of America  
; ZIP: 48226-4415

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS v.6.22  
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/303,025  
; FILING DATE: 08-SEPT-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06829  
; FILING DATE: 14-AUG-1992  
; APPLICATION NUMBER: US 08/152,488  
; FILING DATE: 12-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 313-496-7622  
; TELEFAX: 313-496-8454

INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-303-025-15

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKKAACKAKKAACKAKKAACKAKKAACK 32

## RESULT 3

US-08-677-304-13  
; Sequence 13, Application US/08677304  
; Patent No. 5721212  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 512 Springfield Avenue  
; CITY: Cranford  
; STATE: New Jersey  
; COUNTRY: United States of America  
; ZIP: 07016-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/677,304  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/152,488  
; FILING DATE: 12-NOV-1993  
; APPLICATION NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: 28,664  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-276-3344  
; TELEFAX: 908-276-5543  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: No. 5721212 Relevant  
; TOPOLOGY: No. 5721212 Relevant  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
US-08-677-304-13

Query Match 40.7%; Score 67.5; DB 1; Length 32;  
Best Local Similarity 61.3%; Pred. No. 0.05;  
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKYAKK-EKAACKAYKKEAKAKAAEAAAKE 30  
||| ||| :||||| | ||| ||| :  
Db 2 AKKAACKAKKAACKAKKAACKAKKAACK 32

## RESULT 4

US-08-436-703B-2



```
; Sequence 2, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 6601 Woodward Avenue
; STREET: Suite 1525
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk 1.44Mb, 3.5"
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6;
; SOFTWARE: ASCII (DOS)text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,703B
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: N/A
; FILING DATE: N/A
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: 7WK-060548-00233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-965-1976
; TELEFAX: 313-965-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; US-08-436-703B-2

Query Match 40.7%; Score 67.5; DB 2; Length 32;
Best Local Similarity 61.3%; Pred. No. 0.05;
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKYAKK-EKAAKKAYKKEAKAKAAEAAAKE 30
| | | | | : | | | | | | | | | | | |
Db 2 AKAAKAKAKAAKAAKAAKAAKAAKAAK 32

RESULT 5
US-08-303-025-16
; Sequence 16, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
```

```
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS V.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-7622
; TELEFAX: 313-496-8454
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-16

Query Match 40.7%; Score 67.5; DB 1; Length 33;
Best Local Similarity 61.3%; Pred. No. 0.051;
Matches 19; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 1 AKYAKK-EKAAKKAYKKEAKAKAAEAAAKE 30
| | | | | : | | | | | | | | | | | |
Db 3 AKAAKAKAKAAKAAKAAKAAKAAKAAK 33

RESULT 6
US-08-436-703B-4
; Sequence 4, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 6601 Woodward Avenue
; STREET: Suite 1525
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
```



```
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,488
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: PCT/US92/08069
; APPLICATION NUMBER: 14-AUG-1993
; FILING DATE: 14-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: RM-7WG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-276-3344
; TELEFAX: 908-276-5543
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-152-488-10
Query Match 37.7%; Score 62.5; DB 1; Length 29;
Best Local Similarity 62.1%; Pred.No. 0.17;
Matches 18; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
QY 1 AKKYAKKEKAAYKYEAAAEAAA 29
DB 2 AKKAACKAATKAKKAAPAK-KAKKAATKA 29
RESULT 11
US-08-152-488-11
; Sequence 11, Application US/08152488
; Patent No. 5534619
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 512 Springfield Avenue
; CITY: Cranford
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07016-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,488
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: PCT/US92/08069
; APPLICATION NUMBER: 14-AUG-1993
; FILING DATE: 14-AUG-1993
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: RM-7WG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-276-3344
; TELEFAX: 908-276-5543
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-152-488-11

Query Match 37.7%; Score 62.5; DB 1; Length 29;
Best Local Similarity 62.1%; Pred. No. 0.17;
Matches 18; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAAYKAYKKAARAAK 29
   ||| ||| ||| ||| ||| ||| |||
Db 2 AKKAACKAKKAAYKKAARAAK 29

RESULT 12
US-08-303-025-10
; Sequence 10, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-8454
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
```

```
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: RM-7WG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-276-3344
; TELEFAX: 908-276-5543
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-10

Query Match 37.7%; Score 62.5; DB 1; Length 29;
Best Local Similarity 62.1%; Pred. No. 0.17;
Matches 18; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAAYKAYKKAARAAK 29
   ||| ||| ||| ||| ||| ||| |||
Db 2 AKKAACKAKKAAYKKAARAAK 29

RESULT 13
US-08-303-025-11
; Sequence 11, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-8454
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 13.1962 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-2  
Perfect score: 213  
Sequence: 1 AKYAKAKAEKAKAYKAA.....AKYEKAAAEKAAAEKAAEAYEA 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	48.6	210	2 A25550	histone H1 - sea u
2	98	46.0	394	2 F90725	membrane spanning
3	98	46.0	394	2 G85576	membrane spanning
4	98	46.0	421	2 JV0057	colA protein - Esc
5	94.5	44.4	239	2 A61317	hypothetical prote
6	90	42.3	243	2 AE1689	hypothetical prote
7	90	42.3	1390	2 S51364	sperm tail-specifi
8	89.5	42.0	211	2 A28100	histone H1-beta, e
9	89.5	42.0	214	2 G70673	probable hupB - My
10	88	41.3	899	2 B87553	DNA topoisomerase
11	86.5	40.6	384	2 B43592	outer membrane pro
12	86	40.4	199	2 F70742	hypothetical prote
13	86	40.4	344	2 S34153	mat101-1 protein -
14	86	40.4	347	2 E83525	TolA protein PA097
15	85	39.9	182	2 S61926	histone H1 homolog
16	82	38.5	248	1 HSUR1P	histone H1, gonada
17	82	38.5	262	2 T35503	1-acylglycerol-3-p
18	82	38.5	568	2 T39675	asparaginyl-trna s
19	81.5	38.3	311	2 T17698	hypothetical prote
20	80.5	37.8	277	2 T34625	probable NLP/P60 f
21	80.5	37.8	445	2 T50972	probable zootin [i
22	80.5	37.8	952	2 T36664	probable DNA topoi
23	80	37.6	194	1 HSTR1	histone H1 - trout
24	80	37.6	206	1 HSTR1	histone H1 - rainb
25	80	37.6	291	2 AB1487	hypothetical prote
26	79.5	37.3	217	2 A26721	histone H1-gamma,
27	79	37.1	62	2 T30977	hypothetical prote
28	79	37.1	205	2 S19114	cgr-1 protein - C
29	79	37.1	244	2 S40436	histone H1 - midge

#### RESULT 1

A25550

histone H1 - sea urchin (Lytechinus pictus)

C;Species: Lytechinus pictus (painted urchin)

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999

C;Accession: A25550

R;Knowles, J.A.; Childs, G.J.

Nucleic Acids Res. 14, 8121-8133, 1986

A;Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus and

A;Reference number: A25550; MUID:87040778; PMID:3022245

A;Accession: A25550

A;Molecule type: DNA

A;Residues: 1-210 <KNO>

A;Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617

C;Superfamily: histone H1

C;Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 48.6%; Score 103.5; DB 2; Length 210;

Best Local Similarity 65.1%; Pred. No. 0.0053;

Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

Qy 1 AKYAKAKAEK-AKYAKAAEAKKAAKYEKAAAEKAAAEKAA 42

Db 166 AKKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 208

#### RESULT 2

F90725

membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain 1

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C;Accession: F90725

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: F90725

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-394 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA834197.1; PID:gl33360233; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: EC80774

Query Match 46.0%; Score 98; DB 2; Length 394;

Best Local Similarity 65.3%; Pred. No. 0.028;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

Qy 2 KKYAKKAKAEKAKAEKAKAEKAKAEKAKAEKAKAEKAKAEKAA 45

Db 166 AKKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 208

Db 220 KKAEEKAAAEKAAADKKAEEKAAADKKAEEKAAAEKAAAEKAAAEKAAAE 267

RESULT 3

G85576

membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Nov-2001

C;Accession: G85576

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: G85576

A;Status: Preliminary

A;Molecule type: DNA

A;Residues: 1-394 <STO>

A;Cross-references: GB:AE005174; NID:g12513672; PIDN:AAG55075.1; GSPDB:GN00145; UMGP:209

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: tolA

Query Match 46.0%; Score 98; DB 2; Length 394;

Best Local Similarity 65.3%; Pred. No. 0.028;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAAAEKA---KK--AYKAAEAKKAAKYKAAAEKAAAEKAAAEKAAAE 45

Db 220 KKAEEKAAAEKAAADKKAEEKAAADKKAEEKAAAEKAAAEKAAAEKAAAE 267

RESULT 4

JV0057

tolA protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002

C;Accession: JV0057; B64810

R;Levengood, S.K.; Webster, R.E.

J. Bacteriol. 171, 6600-6609, 1989

A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their products

A;Reference number: JV0057; MUID:90078104; PMID:2687247

A;Accession: JV0057

A;Molecule type: DNA

A;Residues: 1-421 <LEV>

A;Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019

A;Experimental source: strain JM105

A;Note: the authors translated the initiation codon GTG for residue 1 as Val

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64810

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <BLAT>

A;Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73833.1; PID:g1786960;

A;Experimental source: strain K-12, substrain MG1655

C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach the cell surface

C;Genetics:

A;Gene: tolA

A;Map position: 17 min

A;Start codon: GTG

C;Keywords: nucleotide binding; P-loop; transmembrane protein

F;14-34/Domain: transmembrane #status predicted <MSS>

F;78-301/Domain: helical #status predicted <HSG>

F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 46.0%; Score 98; DB 2; Length 421;

Best Local Similarity 65.3%; Pred. No. 0.029;

Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAAAEKA---KK--AYKAAEAKKAAKYKAAAEKAAAEKAAAEKAAAE 45

Db 247 KKAEEKAAAEKAAADKKAEEKAAADKKAEEKAAAEKAAAEKAAAEKAAAE 294

RESULT 5

AE1317

hypothetical protein lmo1941 [imported] - Listeria monocytogenes (strain EGD-e)

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AE1317

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. Science 294, 849-852, 2001

D.; Jones, L.M.; Karst, U.

A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AE1317

A;Status: Preliminary

A;Molecule type: DNA

A;Residues: 1-239 <GLA>

A;Cross-references: GB:NC 003210; PIDN:CAD00019.1; PID:g16411394; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo1941

Query Match 44.4%; Score 94.5; DB 2; Length 239;

Best Local Similarity 56.2%; Pred. No. 0.04;

Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;

QY 2 KKYAKKAAAEKAKK---AYKAAEAKK--AAKYKAAAEKAAAEKAAAE 44

Db 124 KAAAEKAAEDKKQBEDAVKAAAKKEQEAEEKAAADKAAAEKAAAE 171

RESULT 6

AE1689

hypothetical protein homolog lin2055 [imported] - Listeria innocua (strain Clp11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AE1689

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. Science 294, 849-852, 2001

D.; Jones, L.M.; Karst, U.

A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AE1689

A;Status: Preliminary

A;Molecule type: DNA

A;Residues: 1-243 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC97285.1; PID:g16414556; GSPDB:GN00178

A;Experimental source: strain Clp11262

C;Genetics:

A;Gene: lin2055

Query Match 42.3%; Score 90; DB 2; Length 243;

Best Local Similarity 45.5%; Pred. No. 0.11;

Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;

QY 1 AKKYA-KYKAAEKAKYKAAAEKKA-----AKYEKAAAEKAAA 38

Db 106 AKKAAEEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAE 165

QY 39 KEAAE 44

Db 166 DKAAE 171



probable hupB - Mycobacterium tuberculosis (strain H37RV)  
C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C/Accession: G70673  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.











```

RESULT 8
HBHA MYCTU
ID HBHA MYCTU STANDARD; PRT; 198 AA.
AC Q11142; O85733;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heparin-binding hemagglutinin (Adhesin).
GN HBHA OR RV0475 OR MTO493 OR MTCY20G9.01.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=H37RV, ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=98445421; PubMed=9770536;
RA Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Locht C.;
RT "Molecular characterization of the mycobacterial heparin-binding
RT hemagglutinin, a mycobacterial adhesin.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12625-12630(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Barry C.E. III, Tekala F.,
RA Gordon S.V., Eiglmeier K., Gae S., Garsy C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-16, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=ATCC 201 / H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=97188915; PubMed=9064359;
RA Menozzi F.D., Rouse J.H., Alavi M., Laude-Sharp M., Muller J.,
RA Bischoff R., Brennan M.J., Locht C.;
RT "Identification of a heparin-binding hemagglutinin present in
RT Mycobacteria.";
RL J. Exp. Med. 184:993-1001(1996).
RN [5]
RP FUNCTION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=103, and BCG;
RX MEDLINE=21342355; PubMed=11449276;
RA Pethé K., Alonso S., Biet F., Delogu G., Brennan M.J., Locht C.,
RA Menozzi F.D.;
RT "The heparin-binding haemagglutinin of M. tuberculosis is required for
RT extrapulmonary dissemination.";
RL Nature 412:190-194(2001).
CC -1- FUNCTION: REQUIRED FOR EXTRAPULMONARY DISSEMINATION. MEDIATES
CC ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED
CC GLYCOCONGUATES PRESENT AT THE SURFACE OF THESE CELLS; BINDS
CC HEPARIN, DEXTRAN SULFATE, FUCOIDAN AND CHONDROITIN SULFATE.

```

```

CC PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.
CC INDUCES MYCOBACTERIAL AGGREGATION.
CC -1- SUBCELLULAR LOCATION: SURFACE ASSOCIATED.
CC -1- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL
CC DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END
CC DIMINISH THE AFFINITY FOR HEPARIN.
CC -1- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
CC PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
CC SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
CC TERMINAL DOMAIN OF HBHA.
CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
CC RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CC CONTAIN ANY.
CC -1- SIMILARITY: STRONG, TO M.LEPRAE HBHA.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF074390; AAC26052.1; -.
CC EMBL; Z77162; CAB00936.1; -.
CC TIGR; MTO493; -.
CC Tuberculin; RV0475; -.
CC Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
CC Virulence; Complete proteome.
CC INIT MET 0
CC DOMAIN 151 193 ALA/LYS-RICH.
CC CONFLICT 120 120 R -> P (IN REF. 1).
CC SEQUENCE 198 AA; 21403 MW; 513760F6F1EB6042 CRC64;
CC
CC Query Match 40.4%; Score 86; DB 1; Length 198;
CC Best Local Similarity 65.6%; Pred. No. 0.068;
CC Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
CC
CC QY 11 EKAKYKAAEAKKAAKYEKAAEAKKAAEAKA 42
CC DB 160 KKAAPAKKAAKAAKAAKAAKAAKAAKAA 191
CC
CC RESULT 9
CC MSLT_DROHY
CC ID MSLT_DROHY STANDARD; PRT; 344 AA.
CC AC Q08695;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Axoneme-associated protein mst101(1).
CC GN MSL101(1).
CC OS Drosophila hydei (Fruit fly).
CC OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
CC OX NCBI_TaxID=7224;
CC RN [1]
CC RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
CC RC TISSUE=Testis;
CC RX MEDLINE=94200512; PubMed=8150205;
CC RA Neesen J., Buenemann H., Heinlein U.A.;
CC RT "The Drosophila hydei gene hmst101(1) encodes a testis-specific,
CC repetitive, axoneme-associated protein with differential abundance in
CC Y chromosome deletion mutant flies.";
CC RL Dev. Biol. 162:414-425(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC SPERMATID BUNDLES.

```

CC -I- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
 CC -I- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; X73480; CAAS1875.1; -.  
 DR PIR; S34153; S34153.  
 DR FlyBase; FBgn0011816; Dhyd\met101(1).  
 KW SpERM; Repeat; Multigene family.  
 FT DOMAIN 58 337 19 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT K-K-K-X-X-E-X-A-[KQ]-K-X-X-E-X-A-X.  
 SQ SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 344;  
 Best Local Similarity 57.8%; Pred. No. 0.11;  
 Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;  
 QY 2 KKYAKKAK-AEKKAKYKAAEAKK---AAKYKAAAEKAAAEKAA 42  
 DB 117 KEAAEKKKCAEAAKEAAEKKKCAEAAKEAAEKKKCAEAA 161  
 RESULT 10  
 TOLA\_PSEAE STANDARD; PRT; 347 AA.  
 AC P50600;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ToIA protein.  
 GN TOLA OR PA0971.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas  
 OC NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAO;  
 RA Dennis J.J., Lafontaine E.R., Sokol P.A.;  
 RX MEDLINE=97113525; PubMed=8955385;  
 RT "Identification and characterization of the tolQRA genes of  
 RT Pseudomonas aeruginosa."  
 RL J. Bacteriol. 178:7059-7068(1996).  
 RN [2]  
 RP REVISIONS TO N-TERMINUS.  
 RA Duan K., Sokol P.A.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAO1;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagie W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Goltzy S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
 RT opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 CC -I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS  
 CC (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
 CC (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; U39558; AAC44660.2; -.  
 DR EMBL; AE004530; AAG04360.1; -.  
 KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;  
 KW Complete proteome.  
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 17 37 POTENTIAL.  
 FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 209 216 POLY-ALA.  
 SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;  
 Query Match 40.4%; Score 86; DB 1; Length 347;  
 Best Local Similarity 51.1%; Pred. No. 0.11;  
 Matches 24; Conservative 6; Mismatches 15; Indels 2; Gaps 1;  
 QY 1 AKKYAKKAKAEKAAKKA--YKAAEAKKAAKYKAAAEKAAAEKAAAE 45  
 DB 125 AOKAAEAKKADAEKAAEAKKAAEAKKAAEAKKAAEAKKAAE 171  
 RESULT 11  
 DBH\_MYCBO STANDARD; PRT; 205 AA.  
 AC Q9XB18; Q9S5J5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).  
 GN HUP OR HLP OR MDP1.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AN5;  
 RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
 RX "HLPmt-A target for differentiation of M.tuberculosis and M.bovis."  
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BCG / Tokyo;  
 RA Matsumoto S., Yukitake H., Matsuo T., Minoda T., Yamada T.;  
 RT "Identification of a novel protein generating bacterial slow growth  
 RT from Mycobacterium."  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF  
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING  
 CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME  
 CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; Y18421; CAB46493.1; -.  
 DR EMBL; AB013441; BAA78330.1; -.  
 DR HSP; P02346; 1HUU.  
 DR InterPro; IPR000119; Bac\_DNABind.  
 DR InterPro; IPR001386; Histone\_H1/HS.  
 DR Pfam; PF00216; Bac\_DNA\_binding; 1.  
 DR PRINTS; PR00624; HISTONEHS.  
 DR ProDom; PD000945; Bac\_DNABind; 1.



```

DR SMART, SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
KW DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
FT CONFLICT 199 199 A -> T (IN REF. 2).
SQ SEQUENCE 205 AA; 21262 MW; 19FC67885DFE6A8 CRC64;

Query Match 40.1%; Score 85.5; DB 1; Length 205;
Best Local Similarity 58.5%; Pred. No. 0.078;
Matches 24; Conservative 2; Mismatches 12; Indels 3; Gaps 1;

Qy 1 AKYAKAKAKAKAYKAAEAKKAAKYEKAAEAKAAKEA 41
Db 111 AKKVAKKAPAKKATKAAKAAATKAPA---KKAATKAPAKKA 148

RESULT 12
H1_PARAN STANDARD; PRT; 248 AA.
AC P02256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
OS Parachinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinidae;
OC Parachinus.
OX NCBI_TaxID=7658;
RN [1]
RP SEQUENCE OF 1-84.
RA MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N.; Strickland M.; de Groot P.C.; von Holt C.;
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides."
RL Eur. J. Biochem. 104:559-566(1980).
RN [2]
RP SEQUENCE OF 80-248.
RA MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N.; Strickland M.; Brandt W.F.; von Holt C.; Lehmann A.;
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure."
RL Eur. J. Biochem. 104:567-578(1980).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR; A02586; HSUR1P.
DR HSP; P02259; 1HST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
FT VARIANT 144 144 K -> R.
SQ SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 38.5%; Score 82; DB 1; Length 248;
Best Local Similarity 57.8%; Pred. No. 0.2;
Matches 26; Conservative 3; Mismatches 14; Indels 2; Gaps 2;

Qy 2 KKYAKAKAEKAKYKAAEAKKAAKYEKAAEAKAAEAKAA 45
Db 120 KKAATKSAAKAKKA-KAAAKAKAKAKAAAKRRAALAKKAKAAA 163

```

```

RESULT 13
H1E_CHIPA STANDARD; PRT; 235 AA.
AC P40362;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1E.
OS Chironomus pallidivittatus (Widge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7151;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulze E.; Wisniewski J.R.; Nagel S.; Gavenis K.; Grossbach U.;
RA Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L29106; AAA21714.1; -.
CC HSP; P08287; 1GHC.
CC InterPro; IPR001386; Histone H1/H5.
CC InterPro; IPR003216; Linkerhist_N.
CC Pfam; PF00538; linker histone; 1.
CC ProDom; PD000373; Linkerhist_N; 1.
CC SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 235 AA; 24446 MW; 340BC5B9A85002AC CRC64;

Query Match 38.0%; Score 81; DB 1; Length 235;
Best Local Similarity 41.3%; Pred. No. 0.23;
Matches 26; Conservative 2; Mismatches 17; Indels 18; Gaps 1;

Qy 1 AKYAKAKAAEAKKAAKYEKAAEAKAAEAKAAEAKAA 42
Db 126 AKKVDKPKKAPAKPKPKSTNKRVTGKVVKKPAKKATKAAKPAKKVAAKPA 185

Qy 43 YEA 45
Db 186 KKA 188

RESULT 14
TOP1_STRCO STANDARD; PRT; 952 AA.
AC Q9X909;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase I (BC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Untwisting enzyme) (Swivelase).
GN TOPA OR SCO3543 OR SCH5.06C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 23.8291 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989a-2

Perfect score: 213

Sequence: 1 AKYAKKAKAKAKAYKAA.....AKYKAAAKAKAAKAAAYEA 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_podent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	98	46.0	394	16 Q8X965	Q8X965 escherichia
2	94.5	44.4	239	16 Q8Y5W4	Q8Y5W4 listeria mo
3	91	42.7	244	16 Q9AJX2	Q9AJX2 streptomyce
4	88.5	42.3	243	16 Q92A67	Q92A67 listeria in
5	88.5	41.5	275	5 Q01395	Q01395 drosophila
6	88.5	41.5	372	2 Q9WXX1	Q9WXX1 pseudomonas
7	88	41.3	899	16 Q9A5J6	Q9A5J6 caulobacter
8	86.5	40.6	200	16 Q8XVN7	Q8XVN7 raistonia s
9	86	40.4	212	3 Q93946	Q93946 candida alb
10	85	39.9	98	5 Q8WQ44	Q8WQ44 leishmania
11	85	39.9	101	5 Q9BMV8	Q9BMV8 leishmania
12	85	39.9	111	5 Q8T9R3	Q8T9R3 leishmania
13	85	39.9	182	2 Q45370	Q45370 bordetella
14	84	39.4	60	5 Q9U3W3	Q9U3W3 caenorhabdi
15	83.5	39.2	483	12 Q8QND1	Q8QND1 ectocarpus
16	83	39.0	293	10 Q9AT18	Q9AT18 lens culina

17	83	39.0	389	16 Q9CM70	Q9CM70 pasteurella
18	82.5	38.7	243	5 Q23784	Q23784 chironomus
19	82	38.5	262	16 Q9S2M2	Q9S2M2 streptomyce
20	82	38.5	568	3 Q94567	Q94567 schizosacch
21	81.5	38.3	311	12 Q84528	Q84528 paramecium
22	81.5	38.3	482	2 Q93LK4	Q93LK4 enterococcu
23	81	38.0	81	5 Q9NFP6	Q9NFP6 trypanosoma
24	81	38.0	112	5 Q9XYV5	Q9XYV5 leishmania
25	81	38.0	241	5 Q23790	Q23790 chironomus
26	81	38.0	290	10 Q9AT24	Q9AT24 pisum sativ
27	81	38.0	295	10 Q9ZB20	Q9ZB20 pisum sativ
28	81	38.0	295	10 Q9AT22	Q9AT22 lathyrus sa
29	81	38.0	296	10 Q9AT25	Q9AT25 pisum sativ
30	81	38.0	297	10 Q9SXQ8	Q9SXQ8 pisum sativ
31	81	38.0	301	10 Q9AT23	Q9AT23 pisum sativ
32	81	38.0	306	10 Q9AT21	Q9AT21 lathyrus sa
33	80.5	37.8	277	16 Q9XAQ3	Q9XAQ3 streptomyce
34	80.5	37.8	445	3 Q9P3Q8	Q9P3Q8 neurospora
35	80	37.6	66	5 Q9S0Z0	Q9S0Z0 caenorhabdi
36	80	37.6	71	5 Q9NFX8	Q9NFX8 trypanosoma
37	80	37.6	81	5 Q9N6L9	Q9N6L9 trypanosoma
38	80	37.6	281	10 Q9AT20	Q9AT20 lens culina
39	80	37.6	291	16 Q9ZEM5	Q9ZEM5 listeria in
40	80	37.6	293	10 Q9AT19	Q9AT19 lens culina
41	80	37.6	347	16 Q9RKL9	Q9RKL9 streptomyce
42	80	37.6	969	5 Q9NDI9	Q9NDI9 plasmodium
43	80	37.6	1340	16 Q9L1H8	Q9L1H8 streptomyce
44	79.5	37.3	407	16 Q8ZQT6	Q8ZQT6 salmonella
45	79	37.1	62	5 Q17536	Q17536 caenorhabdi

ALIGNMENTS

RESULT 1

Q8X965	PRELIMINARY;	PRT;	394 AA.
ID Q8X965			
AC Q8X965;			
DT 01-MAR-2002 (TrEMBLrel. 20, Created)			
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE Membrane spanning protein, required for outer membrane integrity			
DE (Membrane spanning protein Tola)			
GN TOLA OR Z0907 OR ECS0774.			
OS Escherichia coli O157:H7.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia.			
OX NCBI_TaxID=83334;			
RN [1]			
RC SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX MEDLINE=21074935; PubMed=11206551;			
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,			
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA Welch R.A., Blattner F.R.;			
RT "genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL Nature 409:529-533(2001).			
RN [2]			
RC SEQUENCE FROM N.A.			
RC STRAIN=O157:H7 / RIMD 0509952;			
RX MEDLINE=21156231; PubMed=11258796;			
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL DNA Res. 8:11-22(2001).			
DR EMBL; AB005252; AAG55075.1; -.			
DR EMBL; AF002553; BAB34197.1; -.			

```

DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PRO0308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5B58D8E8230BDE28 CRC64;

Query Match 46.0%; Score 98; DB 16; Length 394;
Best Local Similarity 65.3%; Pred. No. 0.04;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

Qy 2 KKYAKKAKAKA---KK--AYKAAEKAAKAYKAAEKAAEKAAEKAAEAAVEA 45
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 220 KAAEKAAEKAAADKKAAEKAAADKKAA--AKAAEKAAEKAAEKAAEAA 267

RESULT 2
Q8Y5W4 PRELIMINARY; PRT; 239 AA.
ID Q8Y5W4
AC Q8Y5W4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo1941.
GN LMO1941.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kluft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek B., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591981; CAD00019.1; -.
DR ListiList; LMO01941; -.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;

Query Match 44.4%; Score 94.5; DB 16; Length 239;
Best Local Similarity 56.2%; Pred. No. 0.053;
Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;

Qy 2 KKYAKKAKAKAKK---AYKAAEK---AAKYKAAEKAAEKAAEKAAEAAVE 44
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 124 KAAEKAAEKKKQEDAVKAAAKKEQEAEEKAAADKKAAEKAAEKAAEAA 171

RESULT 3
Q9AJX2 PRELIMINARY; PRT; 244 AA.
ID Q9AJX2
AC Q9AJX2
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted protein.
GN SCO1805 OR SCI33.04.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

```

---

```

OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL513407; CAC28545.1; -.
SQ SEQUENCE 244 AA; 25524 MW; 61999D62CA23A7B0 CRC64;

Query Match 42.7%; Score 91; DB 16; Length 244;
Best Local Similarity 59.5%; Pred. No. 0.12;
Matches 25; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

Qy 5 AKKAAEKAKK---AYKAAEKAAKAYKAAEKAAEKAAEKAAEAAVE 44
::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 78 SQKAAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAE 119

RESULT 4
Q92A67 PRELIMINARY; PRT; 243 AA.
ID Q92A67
AC Q92A67
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein lin2055.
GN LIN2055.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kluft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek B., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97285.1; -.
DR ListiList; LIN02055; -.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 25963 MW; 6B2493D143B159D1 CRC64;

Query Match 42.3%; Score 90; DB 16; Length 243;
Best Local Similarity 45.5%; Pred. No. 0.15;
Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;

Qy 1 AKKYA-KKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAE 38
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 106 AKKAAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAEKAAE 165

```

```
QY 39 KRAVE 44
Db 166 DKAKE 171

RESULT 5
O01395 PRELIMINARY; PRT; 275 AA.
AC O01395;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Axoneme-associated protein MST101(3).
GN MST101(3) OR DHMST101.
OS Drosophila hydei (fruit fly).
OC Rukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Periygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]

SEQUENCE FROM N.A.
RA Naesen J., Heinlein U.A.O., Buensmann H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -
DR FlyBase; FBgn020732; Dhyd\mat101(3).
KW Sperm; Repeat; Multigene family.
FT DOMAIN 64 255
FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-X-
FT [AE]-X.
SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2A2DF732C CRC64;

Query Match 41.5%; Score 88.5; DB 5; Length 275;
Best Local Similarity 61.0%; Pred. No. 0.23;
Matches 25; Conservative 1; Mismatches 12; Indels 3; Gaps 1;

QY 5 AKKAAKAKKAYKAAKAKK--AAKYKAAKAAKAAKAA 42
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 AKKCAAAKKEKAAKAAKAAKAAKAAKAAKAAKAA 103

RESULT 6
Q9WWX1 PRELIMINARY; PRT; 372 AA.
AC Q9WWX1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Tola protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]

SEQUENCE FROM N.A.
RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RA Rodriguez-Herva J.J.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]

RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX MEDLINE=96422022; PubMed=8924639;
RA Rodriguez-Herva J.J., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RL J. Bacteriol. 178:5836-5840(1996).
DR EMBL; X74218; CAB50780.1; -
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 41.5%; Score 88.5; DB 2; Length 372;
Best Local Similarity 39.7%; Pred. No. 0.31;
Matches 27; Conservative 8; Mismatches 10; Indels 23; Gaps 1;

QY 1 AKKYAKKAAKAKKAYKAAKAAKAAKAAKAAKAA 37
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AEDAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 180

QY 38 AKAAKAA 45
||| ||| |||
Db 181 AEAACKKA 188

RESULT 7
Q9ASJ6 PRELIMINARY; PRT; 899 AA.
AC Q9ASJ6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DNA topoisomerase (EC 5.99.1.2).
GN CC2451.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]

SEQUENCE FROM N.A.
RA STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Newton A., Stephens C., Phadke N.D., Haft D.H.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
DR EMBL; AE005914; AAK24422.1; -
DR HSSP; P06612; 1BCL.
DR TIGR; CC2451; -
DR InterPro; IPR002936; DNAPrim_topr.
DR InterPro; IPR003601; DNATopi_ATP_bind.
DR InterPro; IPR003602; DNATopi_DNA_bind.
DR InterPro; IPR000380; Prok_topisomase.
```



```

RESULT 14
Q9U3W3 ID Q9U3W3 PRELIMINARY; PRT; 60 AA.
AC Q9U3W3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Histone H1.Q (Hypothetical 6.2 kDa protein).
GN C01B10.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Jedrusik M.A., Schulze E.;
RT "The histone H1 complement of Caenorhabditis elegans.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Blanchard M., Bradshaw H.;
RT "The sequence of C. elegans cosmid C01B10.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF216291; AAF23175.1; -.
DR EMBL; U58757; AAK66021.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PR00624; HISTONEH5.
KW Hypothetical protein.
SQ SEQUENCE 60 AA; 6212 MW; D01ABB4CEC35566D CRC64;

Query Match 39.4%; Score 84; DB 5; Length 60;
Best Local Similarity 58.1%; Pred. No. 0.14;
Matches 25; Conservative 4; Mismatches 12; Indels 2; Gaps 2;

Qy 2 KYAKKAKAKAKKAY-KAAEAKKAAK-YEKAAEKAARKEAA 42
Db 14 KKVAAPKAPVKVKKASPKAAAPRAKPKVKKAAAKSPAKKAA 56
|||||
|||||

RESULT 15
Q8QND1 ID Q8QND1 PRELIMINARY; PRT; 483 AA.
AC Q8QND1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE EAV-1-151.
OS Ectocarpus siliculosus virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.
OX NCBI_TaxID=37665;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ESV-1;
RC Delaroque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;
RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus
genome.";

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 23.4494 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989a-2

Perfect score: 213

Sequence: 1 AKYAKAKAKAKAYKAA.....AKYKAAAKKAAKAAAYEA 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	45	AAV82572	Copolymer molecule
2	138	64.8	109	AAV82577	Copolymer molecule
3	134.5	63.1	56	AAV82575	Copolymer molecule
4	131	61.5	77	AAV82575	Copolymer molecule
5	126.5	59.4	86	AAV82576	Copolymer molecule
6	120.5	56.6	66	AAV82574	Copolymer molecule
7	108	50.7	35	AAV82571	Copolymer molecule
8	99.5	46.7	106	AAV82571	Recombinant copoly
9	98	46.0	46	AAV82871	High affinity macr
10	94.5	44.4	239	ABB49123	Listeria monocytog

11	94	44.1	154	11	AAV82572	Recombinant copoly
12	91	42.7	223	20	AAV14928	Amino acid sequenc
13	89.5	42.0	214	20	AAV34055	M. tuberculosis hi
14	89.5	42.0	214	21	AAV57353	M. tuberculosis hi
15	87	40.8	334	22	ABG28693	Novel human diagno
16	86	40.4	339	19	AAW44934	Mycobacterial hepa
17	86	40.4	198	19	AAW44934	Mycobacterial hepa
18	86	40.4	198	19	AAW44936	Mycobacterial hepa
19	85.5	40.1	205	21	AAV20575	Mycobacterium bovi
20	85	39.9	427	22	AAV70868	C albicans apoptos
21	82	38.5	165	22	ABG91997	C glutamicum prote
22	77.5	36.4	239	22	ABG63276	Drosophila melanog
23	77.5	36.4	2151	22	ABB60086	Drosophila melanog
24	77	36.2	103	20	AAV34067	Histone H1 isoform
25	77	36.2	103	21	AAV57365	Human histone H1.5
26	77	36.2	116	20	AAV34066	Histone H1 isoform
27	77	36.2	116	21	AAV57364	Human histone H1.5
28	77	36.2	158	20	AAV34068	Histone H1 isoform
29	77	36.2	158	21	AAV57366	Human histone H1.5
30	77	36.2	222	18	AAW29476	Human histone H1
31	77	36.2	222	21	AAV34033	Human histone H1
32	77	36.2	222	21	AAV57331	Human histone H1
33	77	36.2	226	20	AAV34060	Human histone H1
34	77	36.2	226	21	AAV57358	Human histone H1
35	76	35.7	157	22	ABB58855	Drosophila melanog
36	75.5	35.4	111	23	ABP00299	Human ORFX protein
37	75.5	35.4	201	23	ABP41465	Human ovarian anti
38	75.5	35.4	218	21	AAV00755	Human secreted pro
39	75.5	35.4	218	21	AAV00759	Human secreted pro
40	75.5	35.4	220	23	AAU76972	Human ribosomal L1
41	75.5	35.4	225	21	AAV00758	Human secreted pro
42	75.5	35.4	265	21	AAV58221	Lung cancer associ
43	75	35.2	160	18	AAW14549	Streptococcus pneu
44	74.5	35.0	32	16	AAV90180	Polycationic poly
45	74.5	35.0	32	18	AAW06686	Protamine-like pep

#### ALIGNMENTS

##### RESULT 1

AAV82572  
ID AAV82572 standard; peptide; 45 AA.

XX AAV82572;

XX AC

DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence 85Q ID NO:2.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; chymotrypsin; haemostatic; antipsoriatic; dermatological;  
KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PA (TEVA-) TEVA PHARM USA INC.

XX Gad A, Lis D;  
XX WPI; 2000-317499/27.  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
XX glatiramer acetate and for treatment and prevention of immune diseases  
XX  
XX Claim 10; Page 14; 72pp; English.  
XX  
XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
XX weight TV-marker polypeptides from the present invention. The present  
XX invention describes polypeptides (I) for determining the molecular  
XX weight of a copolymer (CP), which has an identified molecular weight  
XX and an amino acid composition corresponding to the copolymer. The  
XX polypeptides of the invention are used as molecular weight markers for  
XX glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX used for treating and preventing immune diseases in a mammal. Autoimmune  
XX diseases which may be treated include either cell-mediated or  
XX antibody-mediated diseases. Such diseases include arthritic conditions,  
XX demyelinating diseases and inflammatory conditions, e.g. multiple  
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX lupus erythematosus. Mediated-mediated diseases which can be treated  
XX include host-versus-graft disease, graft-versus-host disease, and  
XX delayed-type hypersensitivity. The polypeptides of the invention have  
XX defined molecular weights and physical properties which are analogous to  
XX glatiramer acetate molecules, which makes them ideal for use as  
XX molecular weight markers.  
XX  
XX Sequence 45 AA;  
XX  
XX Query Match 100.0%; Score 213; DB 21; Length 45;  
XX Best Local Similarity 100.0%; Pred. No. 1.4e-16;  
XX Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 AKKYAKKAAKAAKAYKAAAEAKK--AAKYKAAAEKAAAEKAAAEYEA 45  
DB 1 AKKYAKKAAKAAKAYKAAAEAKKAAKAYKAAAEKAAAEKAAAEYEA 45  
RESULT 2  
AAY82577  
ID AAY82577 standard; peptide; 109 AA.  
XX  
XX AAY82577;  
XX  
XX 28-JUL-2000 (first entry)  
XX  
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:7.  
XX  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
XX Glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
XX antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
XX antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
XX pemphigus vulgaris; systemic lupus erythematosus.  
XX  
XX Unidentified.  
XX  
XX WO200018794-A1.  
XX  
XX 06-APR-2000.  
XX  
XX 24-SEP-1999; 99WO-US22402.

XX 25-SEP-1998; 98US-0101693.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (TEVA-) TEVA PHARM USA INC.  
XX  
XX Gad A, Lis D;  
XX WPI; 2000-317499/27.  
XX  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
XX glatiramer acetate and for treatment and prevention of immune diseases  
XX  
XX Claim 10; Page 14; 72pp; English.  
XX  
XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
XX weight TV-marker polypeptides from the present invention. The present  
XX invention describes polypeptides (I) for determining the molecular  
XX weight of a copolymer (CP), which has an identified molecular weight  
XX and an amino acid composition corresponding to the copolymer. The  
XX polypeptides of the invention are used as molecular weight markers for  
XX glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX used for treating and preventing immune diseases in a mammal. Autoimmune  
XX diseases which may be treated include either cell-mediated or  
XX antibody-mediated diseases. Such diseases include arthritic conditions,  
XX demyelinating diseases and inflammatory conditions, e.g. multiple  
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX lupus erythematosus. Mediated-mediated diseases which can be treated  
XX include host-versus-graft disease, graft-versus-host disease, and  
XX delayed-type hypersensitivity. The polypeptides of the invention have  
XX defined molecular weights and physical properties which are analogous to  
XX glatiramer acetate molecules, which makes them ideal for use as  
XX molecular weight markers.  
XX  
XX Sequence 109 AA;  
XX  
XX Query Match 64.8%; Score 138; DB 21; Length 109;  
XX Best Local Similarity 76.6%; Pred. No. 5.1e-08;  
XX Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;  
XX  
QY 1 AKKYAKKAAKAAKAYKAAAEAKK--AAKYKAAAEKAAAEKAAAEYEA 45  
DB 65 AKKYAKKAAKAE--KKEYAAAEAKKAAKAAKAYKAAAEKAAAEYEA 109  
RESULT 3  
AAY82573  
ID AAY82573 standard; peptide; 56 AA.  
XX  
XX AAY82573;  
XX  
XX 28-JUL-2000 (first entry)  
XX  
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.  
XX  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
XX Glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
XX antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
XX antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
XX pemphigus vulgaris; systemic lupus erythematosus.  
XX  
XX Unidentified.  
XX  
XX OS  
XX





QY 1 AKYAKAKAEKAKAYKAAEAKKAAKYEKAAEAKAAEAKAAEAAVEA 45  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 29 AKYAKAKAEKAE--KREYAAAEAK-----YKAEAAKAAEAAEAAVEA 66

## RESULT 7

AAV82571  
 ID AAV82571 standard; peptide; 35 AA.

XX AC AAV82571;

XX XX 28-JUL-2000 (first entry)

XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:1.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW Glutiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
 KW Diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.

XX OS Unidentified.

XX XX WO200018794-A1.

XX XX 06-APR-2000.

XX XX 24-SEP-1999; 99WO-US22402.

XX XX 25-SEP-1998; 98US-0101693.

XX XX (YEDA ) YEDA RES & DEV CO LTD.

XX XX (TEVA-) TEVA PHARM USA INC.

XX XX Gad A, Lis D;

XX XX WPI; 2000-317499/27.

XX PT Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glutiramer acetate and for treatment and prevention of immune diseases

XX PS Claim 10; Page 14; 72pp; English.

XX CC AAV82571 to AAV82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glutiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions,  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights and physical properties which are analogous to  
 CC glutiramer acetate molecules, which makes them ideal for use as  
 CC molecular weight markers.

XX SQ Sequence 35 AA;

Query Match 50.7%; Score 108; DB 21; Length 35;  
 Best Local Similarity 64.4%; Pred. No. 2.9e-05;  
 Matches 29; Conservative 1; Mismatches 5; Indels 10; Gaps 2;

QY 1 AKYAKAKAEKAKAYKAAEAKKAAKYEKAAEAKAAEAKAAEAAVEA 45  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1 AKYAKKEKA--AKKAYK-----KEAKAKAAEAAEAAEAAVEA 35

## RESULT 8

AAV06446

ID AAV06446 standard; protein; 106 AA.

XX AC AAV06446;

XX XX 03-JAN-1991 (first entry)

XX DE Recombinant copolymer 1-19, myelin basic protein analogue.

XX KW Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;  
 KW immunological activity; autoimmune encephalomyelitis;  
 KW multiple sclerosis;

XX OS Synthetic.

XX XX EP383620-A.

XX XX 22-AUG-1990.

XX XX 16-FEB-1990; 90EP-0301700.

XX XX 07-FEB-1990; 90US-0473845.

XX XX 17-FEB-1989; 89US-0312541.

XX XX (REPL-) REPLIGEN CORP.

XX XX Cook KS;

XX XX WPI; 1990-255848/34.

XX XX N-PSDB; AAQ06446.

XX PT Producing genes encoding random polymers of aminoacid(s) - for  
 PT producing recombinant polypeptide(s) with biological and/or  
 PT immunological activity

XX PS Disclosure; Fig 12; 25pp; English.

XX CC To improve the expression of rCOP-1 polypeptides in E. coli, genes  
 CC coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN  
 CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
 CC express Protein A. The resulting plasmids encode fusion proteins  
 CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
 CC A methionine residue occurs between the Protein A and rCOP-1  
 CC sequences, originating from the 5' linker sequence, in order that  
 CC the COP-1 polypeptide may be cleaved from the fusion protein.  
 CC rCOP-1-19 contains oligonucleotide duplexes encoding the following  
 CC segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA, KEA, and AAA. The  
 CC N-terminal alanine residue is left behind following CNBr cleavage of the  
 CC fusion protein.  
 CC The product prevents or arrests experimental autoimmune  
 CC encephalomyelitis. They are used to prevent, arrest or control a  
 CC demyelinating disorder, e.g. multiple sclerosis. They may also  
 CC be used as additives to hair care products to confer beneficial  
 CC effects on damaged hair or as supplements for diets deficient in  
 CC certain amino acids.  
 CC See also AAQ05664.

XX SQ Sequence 106 AA;

Query Match 46.7%; Score 99.5; DB 11; Length 106;  
 Best Local Similarity 58.3%; Pred. No. 0.00076;  
 Matches 28; Conservative 4; Mismatches 7; Indels 9; Gaps 2;

Qy 3 KYAKKA-KAEKAKYKAAEAKKAAKAEKAAEAKAA-----AKEA 41  
 ID AAR28871 standard; peptide; 46 AA.  
 AC AAR28871;  
 XX 23-MAR-1993 (first entry)  
 DE High affinity macrophage mannose receptor ligand compound #9.  
 XX glycopeptide; mannose; mannosylated; glycosylated; mannose receptor;  
 KW macrophages; monocytes; destroy; cytotoxicity; label; image; alter;  
 KW macrophage processing of antigen; MHC restriction; inflammation;  
 KW inflammatory diseases; macrophage secretory products;  
 KW Crohn's disease; Legionnaires disease; mononuclear phagocytes; HIV;  
 KW AIDS; lysosomal storage diseases; Gaucher's disease; asthma;  
 KW alveolar macrophages metastasis; systemic macrophages; deliver;  
 KW antigenic peptides; prevent transplant rejection;  
 KW organ transplantation; antitumour agents; cancer; toxins.  
 XX Synthetic.  
 OS  
 XX  
 XX  
 PH Key Location/Qualifiers  
 FT Modified-site 1  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine. May also have non  
 FT interfering substituents."  
 FT  
 FT Modified-site 4  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 7  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 10  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 13  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 16  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 19  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 22  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 25  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 28  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 31  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 34  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 37  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 40  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT  
 FT Modified-site 43

FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine."  
 FT Modified-site 46  
 FT /note= "opt may have mannose, fucose, glucose or  
 FT N-Ac-glucosamine. May also have non  
 FT interfering substituents."  
 XX  
 PN WO9219248-A.  
 XX  
 XX 12-NOV-1992.  
 PD  
 XX  
 XX 01-MAY-1992; 92WO-US03609.  
 PF  
 XX 03-MAY-1991; 91US-0694983.  
 PR  
 XX (UNIW ) UNIV WASHINGTON.  
 PA  
 XX Stahl PD;  
 PI  
 XX WPI; 1992-398516/48.  
 DR  
 XX New high affinity mannose receptor ligand cpds. - for treating  
 FT diseases mediated by macrophage activity e.g. asthma,  
 FT inflammatory diseases and infectious diseases, e.g. HIV  
 XX  
 PS Claim 3; Page 21; 32pp; English.  
 XX  
 CC This compound represents a glycopeptide effective in inhibiting the  
 CC binding of labelled mannosylated BSA to mannose receptors. Mannose  
 CC receptors are uniquely found on macrophages and not on monocytes.  
 CC Glycopeptides such as this provide a mechanism to target macrophages  
 CC specifically, to image, label, destroy or otherwise alter their  
 CC antigen processing function. In addition they can be conjugated to  
 CC solid supports and used to purify mannose receptors from a variety  
 CC of sources. They are useful in the treatment of inflammatory  
 CC diseases driven by macrophage secretory products eg. Crohn's  
 CC disease; infectious diseases in which macrophages harbour replicating  
 CC infectious agents eg. Legionnaires disease; viral infections  
 CC involving mononuclear phagocytes eg. HIV and lysosomal storage  
 CC diseases, in which macrophages are principally involved eg.  
 CC Gaucher's disease; asthma mediated by alveolar macrophages; and in  
 CC controlling metastasis, mediated by systemic macrophages. The peptides  
 CC can also be used to deliver antigenic peptides as conjugates to a  
 CC macrophage to marshal an immune response; also self peptides to  
 CC prevent tissue transplant rejection.  
 XX  
 SQ Sequence 46 AA;  
 Query Match 46.0%; Score 98; DB 13; Length 46;  
 Best Local Similarity 63.4%; Pred. No. 0.00047;  
 Matches 26; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
 Qy 1 AKYAKKAKAEKAKYKAAEAKKAAKAEKAAEAKAAEAKAAEAKAA 41  
 Db 2 AKAKAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 42  
 RESULT 10  
 ABB49123  
 ID ABB49123 standard; Protein; 239 AA.  
 XX  
 XX ABB49123;  
 XX  
 XX 05-FEB-2002 (first entry)  
 DT  
 XX Listeria monocytogenes protein #1827.  
 DE  
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 KW  
 XX Listeria monocytogenes.  
 OS  
 XX WO200177335-A2.  
 FN

XX PD 18-OCT-2001.  
 XX PF 11-APR-2001; 2001WO-FR01118.  
 XX PR 11-APR-2000; 2000FR-0004629.  
 XX PA (INSP ) INST PASTEUR.  
 XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 XX PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Coseart P;  
 XX PI Daniels J, Coebel W, Krafst J, Kuhn M, Ng E, Vazquez-Boland JA;  
 XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 XX PI Chakraborty T, Donann E, Hain T, Berche P, Charbit A, Durant L;  
 XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 XX PI Maduenio E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J;  
 XX PI Rose M, Voss H;  
 XX DR WPI; 2002-010914/01.  
 XX PS Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 XX PT and prevention of *Listeria* and related bacterial infections, and  
 XX PT related polypeptides -  
 XX PS Claim 6; SEQ ID No 1828; 192pp; French.  
 XX CC The present invention relates to the genome sequence of *Listeria*  
 XX CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
 XX CC it are useful for selecting probes and primers for detecting genes in *L.*  
 XX CC *monocytogenes* and related organisms, and for studying genetic  
 XX CC polymorphisms and other genomes. The present sequence is a protein  
 XX CC encoded by the genome sequence of the present invention. Proteins  
 XX CC expressed from the genome sequence are useful for raising specific  
 XX CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 XX CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 XX CC B12. The genome sequence and proteins encoded by it are also useful for  
 XX CC selecting compounds that regulate gene expression and cell replication  
 XX CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 XX CC sequence and proteins encoded by it are useful in pharmaceutical and  
 XX CC vaccines compositions for the treatment or prevention of infections by *L.*  
 XX CC *monocytogenes* and related organisms.  
 XX CC Note: The sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 239 AA;  
 Query Match 44.4%; Score 94.5; DB 23; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.0061;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 QY 2 KKYAKKAKKAKK--AYKAAEAKK--AAKYKAAAEKAAAEKAAAYE 44  
 Db 124 KAAAEKAEADKKKQFEEDAVKAAAKKQEQEAEKAAADKAAAEKAAAE 171  
 RESULT 11  
 AAR06445  
 ID AAR06445 standard; protein; 154 AA.  
 XX AC AAR06445;  
 XX DT 03-JAN-1991 (first entry)  
 XX DE Recombinant copolymer 1-77, myelin basic protein analogue.  
 XX XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
 XX KW immunological activity; autoimmune encephalomyelitis;  
 XX KW multiple sclerosis;  
 XX OS Synthetic.  
 XX PN EF383620-A.

XX PD 22-AUG-1990.  
 XX PF 16-FEB-1990; 90EP-0301700.  
 XX PR 07-FEB-1990; 90US-0473845.  
 XX PR 17-FEB-1989; 89US-0312541.  
 XX PA (REPL-) REPLIGEN CORP.  
 XX PI Cook KS;  
 XX DR WPI; 1990-255848/34.  
 XX DR N-PSDB; AAQ05664.  
 XX PT Producing genes encoding random polymers of aminoacid(s) - for  
 XX PT producing recombinant polypeptide(s) with biological and/or  
 XX PT immunological activity  
 XX PS Disclosure; Fig 11; 25pp; English.  
 XX CC To improve the expression of rCOP-1 polypeptides in *E. coli*, genes  
 XX CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pSG3-2deltaN  
 XX CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
 XX CC express Protein A. The resulting plasmids encode fusion proteins  
 XX CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
 XX CC A methionine residue occurs between the Protein A and rCOP-1  
 XX CC sequences, originating from the 5' linker sequence, in order that  
 XX CC the COP-1 polypeptide may be cleaved from the fusion protein.  
 XX CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
 XX CC segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue  
 XX CC is left behind following CNBr cleavage of the fusion protein.  
 XX CC The product prevents or arrests experimental autoimmune  
 XX CC encephalomyelitis. They are used to prevent, arrest or control a  
 XX CC demyelinating disorder, e.g. multiple sclerosis. They may also  
 XX CC be used as additives to hair care products to confer beneficial  
 XX CC effects on damaged hair or as supplements for diets deficient in  
 XX CC certain amino acids.  
 XX CC See also AAQ05665.  
 XX SQ Sequence 154 AA;  
 Query Match 44.1%; Score 94; DB 11; Length 154;  
 Best Local Similarity 60.5%; Pred. No. 0.0044;  
 Matches 26; Conservative 2; Mismatches 9; Indels 6; Gaps 2;  
 QY 2 KKYAKKAKKAKKAYKAAEAKKAAKYKAAAEKAAAEKAAAYE 44  
 Db 105 KKYKKKAKKAKYKK--KAKEAEKA----KAAAEAKKAEAYK 141  
 RESULT 12  
 AAY14928  
 ID AAY14928 standard; protein; 223 AA.  
 XX AC AAY14928;  
 XX DT 25-OCT-1999 (first entry)  
 XX DE Amino acid sequence of *M. vaccae* antigen GV-45.  
 XX KW Mycobacterium *vaccae* protein; antigen; T cell activation; cytokine;  
 XX KW dendritic cell maturation; infectious disease; immune disorder; cancer;  
 XX KW respiratory system; mycobacterial infection; allergy; tuberculosis;  
 XX KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;  
 XX KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;  
 XX KW squamous cell carcinoma; melanoma.  
 XX OS Mycobacterium *vaccae*.  
 XX PN WO9932634-A2.  
 XX PD 01-JUL-1999.







THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	213	100.0	45	10	US-09-g16-989A-2	Sequence 2, Appli
2	138	64.8	109	10	US-09-g16-989A-7	Sequence 7, Appli
3	134.5	63.1	56	10	US-09-g16-989A-3	Sequence 3, Appli
4	131	61.5	77	10	US-09-g16-989A-5	Sequence 5, Appli
5	126.5	59.4	86	10	US-09-g16-989A-6	Sequence 6, Appli
6	120.5	56.6	66	10	US-09-g16-989A-4	Sequence 4, Appli
7	108	50.7	35	10	US-09-g16-989A-1	Sequence 1, Appli
8	91	42.7	223	9	US-10-051-643-201	Sequence 201, App
9	82	38.5	165	9	US-09-738-626-5751	Sequence 5751, Ap
10	78	36.6	309	9	US-09-820-843A-24	Sequence 24, Appli
11	75.5	35.4	220	10	US-09-923-304-2	Sequence 2, Appli
12	75.5	35.4	265	10	US-09-925-302-559	Sequence 559, App
13	74	34.7	352	9	US-09-820-843A-23	Sequence 23, Appli
14	74	34.7	356	9	US-09-820-843A-27	Sequence 27, Appli
15	72.5	34.0	372	9	US-09-820-843A-8	Sequence 8, Appli
16	70.5	33.1	619	9	US-09-882-774-1	Sequence 1, Appli
17	70	32.9	641	10	US-09-765-272-160	Sequence 160, App
18	70	32.9	1156	10	US-09-915-242-13187	Sequence 13187, A
19	68	31.9	241	10	US-09-938-803-8	Sequence 8, Appli

```
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1998-09-25
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; TYPE: PRT
; LENGTH: 109
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match      64.8%; Score 138; DB 10; Length 109;
Best Local Similarity 76.6%; Pred. No. 3.6e-08;
Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

Qy 1 AKYAKAKAEKAKYKAEAKK--AAKYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 AKYAKAKAE--KKEYAAAEKAAKAEAKYKAAAEKAAAEKAAAEYEA 109

RESULT 3
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

Query Match      63.1%; Score 134.5; DB 10; Length 56;
Best Local Similarity 68.4%; Pred. No. 4.2e-08;
Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAEKAAKAEKAAKAEKAAKAEKAAKAEKAAAEYEA 56

RESULT 4
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; TYPE: PRT
; LENGTH: 77
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match      61.5%; Score 131; DB 10; Length 77;
Best Local Similarity 50.6%; Pred. No. 1.4e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAEKAAKAEKAAKAEKAAKAEKAAKAEKAAAEYEA 77

RESULT 5
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match      59.4%; Score 126.5; DB 10; Length 86;
Best Local Similarity 45.3%; Pred. No. 4.5e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAEKAAKAEKAAKAEKAAKAEKAAKAEKAAAEYEA 86

RESULT 6
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
```

```
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; TYPE: PRT
; LENGTH: 77
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match      61.5%; Score 131; DB 10; Length 77;
Best Local Similarity 50.6%; Pred. No. 1.4e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAEKAAKAEKAAKAEKAAKAEKAAKAEKAAAEYEA 77

RESULT 5
US-09-816-989A-6
; Sequence 6, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match      59.4%; Score 126.5; DB 10; Length 86;
Best Local Similarity 45.3%; Pred. No. 4.5e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3;

Qy 1 AKYAKK-----AKAEKA-----KKAYKAAAEKAAAEKAAAEYEA 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AKYAKKEKAYKAEKAAKAEKAAKAEKAAKAEKAAKAEKAAAEYEA 86

RESULT 6
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
```

```

; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051.643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

Query Match 42.7%; Score 91; DB 9; Length 223;
Best Local Similarity 59.6%; Pred. No. 0.0059;
Matches 28; Conservative 5; Mismatches 12; Indels 2; Gaps 2;

QY 1 AKKVA-KKAAEKAKKAYKAAEKAAKAAKYEKAA-AEKAAAEKAAVEA 45
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 137 AKKAATKAAPAKKATAKKAAPAKKATAKKAAPAKKATAKKAAPAKKATAKKA 183
      |||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-738-626-5751
; Sequence 5751, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5751
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

Query Match 38.5%; Score 82; DB 9; Length 165;
Best Local Similarity 62.9%; Pred. No. 0.038;
Matches 22; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 7 KAKAEKAKKAYKAAEKAAKAAKYEKAAAEKAAAEKAA 41
      :|||:|||||:|||||:|||||:|||||:|||||:
DB 120 EAITEKKKKAREDKAEKAA--EKAAAEKAAAEAS 152
      :|||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
US-09-820-843A-24
; Sequence 24, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:

```



US-09-820-843A-27

Query Match 34.7%; Score 74; DB 9; Length 356;  
Best Local Similarity 45.5%; Pred. No. 0.57;  
Matches 20; Conservative 4; Mismatches 20; Indels 0; Gaps 0;

**Qy**      2    KKVAKKAAEKAKKAYKAAEAKKAAYEKAAAEKAAAEYEA    45  
         | : | | : : | : | | | | | | | : | |  
**Dd**     195 KEVAAKAEQEERLAKKEKAKEAADKAKEKERAAKAAEAERKAOEA    238

## RESULT 15

US-09-820-843A-8

Sequence 8, Application US/09820843A

; Publication No. US20030039963A1

; GENERAL INFORMATION:

; APPLICANT: Council of Scientific and Industrial Research

RESEARCH, COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN

; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES

; FILE REFERENCE: 063915

FILE REFERENCE: Q05913  
CURRENT APPLICATION NUMBER: US/09/820,843A

; CURRENT FILING DATE: 2001-03-30

NUMBER OF SEO ID NOS: 118

: NUMBER OF SEQ ID NOS: 118  
 : SOFTWARE: PatentIn version 3.0

: SOFTWARE: F  
: SEQ ID NO 8

```

; SEQ ID NO 8
; LENGTH: 372

```

```

; LENGTH: 37
. TYPE: PRT

```

TYPE: PR1  
ORGANISM: H. influenzae

; ORGANISM  
; FEATURE:

```
; FEATURE:
; NAME/KEY: misc feature
```

```

; NAME/KEY: misc feature
; OTHER INFORMATION: outer membrane integrity protein (tolA)

```

```

; OTHER INFORMATION: C
: NAME/KEY: misc feat

```

```
; NAME/KEY: misc_feature
: OTHER_INFORMATION: a111573353
```

; OTHER INFORMATION  
IIS-09-930-843A-8

Query Match	Score 72.5;	DB 9;	Length 372;
34.0%			

Best Local Similarity 44.7%; Pred. No. 0.85;

Sequence similarity	24.0%; 120,000 bases	24.0%; 120,000 bases
Matches	21; Conservative	8; Mismatches
Indels	3; Gaps	1; Gaps

QY 2 KKYAKKAAEKAKKAYKA -- AEAKKAAKYEKAAAEEKAAAEKAAEYEA 45

Search completed: March 10, 2003, 12:53:42

Job time : 13.5289 secs

**THIS PAGE BLANK (USPTO)**



Result No.	Score	Query Match	ID			Description
			DB	Length	ID	
1	91	42.7	223	4	US-09-095-855-201	Sequence 201, Appl
2	91	42.7	223	4	US-09-205-426-201	Sequence 201, App
3	89.5	42.0	214	3	US-09-041-889-27	Sequence 27, Appl
4	77	36.2	103	3	US-09-041-889-39	Sequence 39, Appl
5	77	36.2	116	3	US-09-041-889-38	Sequence 38, Appl
6	77	36.2	158	3	US-09-041-889-40	Sequence 40, Appl
7	77	36.2	222	3	US-09-041-889-3	Sequence 3, Appl
8	77	36.2	222	3	US-08-837-058-3	Sequence 3, Appl
9	77	36.2	226	3	US-09-041-889-32	Sequence 32, Appl
10	74.5	35.0	32	1	US-08-152-488-13	Sequence 13, Appl
11	74.5	35.0	32	1	US-08-303-025-15	Sequence 15, Appl
12	74.5	35.0	32	1	US-08-677-304-13	Sequence 13, Appl
13	74.5	35.0	32	2	US-08-436-703B-2	Sequence 2, Appl
14	74.5	35.0	33	1	US-08-303-025-16	Sequence 16, Appl
15	74.5	35.0	33	2	US-08-436-703B-4	Sequence 4, Appl
16	72	33.8	29	1	US-08-152-488-10	Sequence 10, Appl
17	72	33.8	29	1	US-08-152-488-11	Sequence 11, Appl
18	72	33.8	29	1	US-08-303-025-10	Sequence 10, Appl
19	72	33.8	29	1	US-08-303-025-11	Sequence 11, Appl
20	72	33.8	29	1	US-08-303-025-13	Sequence 13, Appl
21	72	33.8	29	1	US-08-677-304-10	Sequence 10, Appl
22	72	33.8	29	1	US-08-677-304-11	Sequence 11, Appl
23	72	33.8	29	2	US-08-436-703B-3	Sequence 3, Appl
24	72	33.8	29	4	US-08-436-703B-15	Sequence 15, Appl
25	71.5	33.6	49	2	US-08-995-172-14	Sequence 14, Appl
26	71.5	33.6	49	4	US-08-839-624-26	Sequence 26, Appl
27	71.5	33.6	49	4	US-09-150-812-26	Sequence 26, Appl

9. AFFILIATION NUMBER: US/05/041,889

```

Query Match          36.2%; Score 77; DB 3; Length 103;
Best Local Similarity 52.2%; Pred. No. 0.068;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKAVKKTTPKAKKPAA 96

RESULT 5
US-09-041-889-38
; Sequence 38, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-38

; Query Match          36.2%; Score 77; DB 3; Length 116;
; Best Local Similarity 52.2%; Pred. No. 0.077;
; Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKAVKKTTPKAKKPAA 96

RESULT 6
US-09-041-889-40
; Sequence 40, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-38

; Query Match          36.2%; Score 77; DB 3; Length 116;
; Best Local Similarity 52.2%; Pred. No. 0.077;
; Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKAVKKTTPKAKKPAA 96

RESULT 6
US-09-041-889-40
; Sequence 40, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-40

; Query Match          36.2%; Score 77; DB 3; Length 158;
; Best Local Similarity 52.2%; Pred. No. 0.11;
; Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAAEKA-----KKYKAAEAKKAAKYKAAAEKAAA 38
Db 51 AKPKAKGAGAAKAKPKAGATPKKAAKAAAGAKAVKKTTPKAKKPAA 96

RESULT 7
US-09-041-889-3
; Sequence 3, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..222
; OTHER INFORMATION: /note= "product = Human Histone
; OTHER INFORMATION: H1-S-3"
US-09-041-889-3

Query Match 36.2%; Score 77; DB 3; Length 222;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAKEA-----KKAYKAAAEAKKAAKYEKAAAEKAAA 38
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 AKPKAKGAGAAKAKKPAGATPKKAKKAGAKKAVKKTPTPKAKKPAA 163

RESULT 8
US-08-837-058-3
; Sequence 3, Application US/08837058
; Patent No. 6074835
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan R.
; APPLICANT: Targan, Stephan R.
; APPLICANT: Eggena, Mark
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Histone H1
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/837,058
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-32

Query Match 36.2%; Score 77; DB 3; Length 226;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

Qy 1 AKKYAKKAKEA-----KKAYKAAAEAKKAAKYEKAAAEKAAA 38
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 AKPKAKGAGAAKAKKPAGATPKKAKKAGAKKAVKKTPTPKAKKPAA 164

RESULT 10
US-08-152-488-13
; Sequence 13, Application US/08152488
; Patent No. 5534619
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
```

; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 512 Springfield Avenue  
; CITY: Cranford  
; STATE: New Jersey  
; COUNTRY: United States of America  
; ZIP: 07016-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/152,488  
; FILING DATE: 12-NOV-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REGISTRATION NUMBER: 28,664  
; REFERENCE/DOCKET NUMBER: RM-7WG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-276-3344  
; TELEFAX: 908-276-5543  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
; US-08-152-488-13

Query Match 35.0%; Score 74.5; DB 1; Length 32;  
Best Local Similarity 68.8%; Pred. No. 0.038;  
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

QY 5 AKKAKAEKAKYKAAEAKKAAKYEKAAAEKA 36  
||| | : ||||| | : ||||| | : |||  
Db 2 AKKA-AKKA--AKKAKAAKAAKAAKKA 30

RESULT 11  
US-08-303-025-15  
; Sequence 15, Application US/08303025  
; Patent No. 5614494  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 150 West Jefferson, Suite 2500

; CITY: Detroit  
; STATE: Michigan  
; COUNTRY: United States of America  
; ZIP: 48226-4415  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS v.6.22  
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/303,025  
; FILING DATE: 08-SEPT-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06829  
; FILING DATE: 14-AUG-1992  
; APPLICATION NUMBER: US 08/152,488  
; FILING DATE: 12-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 313-496-7622  
; TELEFAX: 313-496-8454  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: N/A  
; PUBLICATION INFORMATION:  
; AUTHORS: N/A  
; TITLE: N/A  
; DOCUMENT NUMBER: PCT/US92/08069  
; FILING DATE: 14-AUG-1993  
; US-08-303-025-15

Query Match 35.0%; Score 74.5; DB 1; Length 32;  
Best Local Similarity 68.8%; Pred. No. 0.038;  
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

QY 5 AKKAKAEKAKYKAAEAKKAAKYEKAAAEKA 36  
||| | : ||||| | : ||||| | : |||  
Db 2 AKKA-AKKA--AKKAKAAKAAKAAKKA 30

RESULT 12  
US-08-677-304-13  
; Sequence 13, Application US/08677304  
; Patent No. 5721212  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 512 Springfield Avenue  
; CITY: Cranford  
; STATE: New Jersey  
; COUNTRY: United States of America  
; ZIP: 07016-1811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 6; ASCII (DOS)Text

RESULT 13  
 US-08-436-703B-2  
 Sequence 2, Application US/08436703B  
 Patent No. 5919761  
 GENERAL INFORMATION:  
 APPLICANT: Wakefield, Thomas W.  
 APPLICANT: Andrews, Philip C.  
 APPLICANT: Stanley, James C.  
 TITLE OF INVENTION: NOVEL PEPTIDES FOR  
 TITILE OF INVENTION: HEPARIN AND LOW MOLECULAR  
 TITILE OF INVENTION: WEIGHT HEPARIN  
 TITILE OF INVENTION: ANTICOAGULATION REVERSAL  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSES: Benita J. Rohm, Esq.  
 STREET: 6601 Woodward Avenue  
 STREET: Suite 1525  
 CITY: Detroit  
 STATE: Michigan  
 COUNTRY: United States of America  
 ZIP: 48226  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk 1.44MB, 3.5"  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WordPerfect 6;  
 SOFTWARE: ASCII (DOS)text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,703B  
 FILING DATE: 08-MAY-1995  
 CLASSIFICATION: 514

RESULT 14  
US-08-303-025-16  
; Sequence 16, Application US/08303025  
; Patent No. 5614494  
; GENERAL INFORMATION:  
; APPLICANT: Wakefield, Thomas W.  
; APPLICANT: Andrews, Philip C.  
; APPLICANT: Stanley, James C.  
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND  
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN  
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benita J, Rohm, Esq.  
; STREET: 150 West Jefferson, Suite 2500  
; CITY: Detroit  
; STATE: Michigan  
; COUNTRY: United States of America  
; ZIP: 48226-4415  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS v.6.22  
; SOFTWARE: Wordperfect 6.1; ASCII (DOS)Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/303,025  
; FILING DATE: 08-SEPT-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06829  
; FILING DATE: 14-AUG-1992  
; APPLICATION NUMBER: US 08/152,488  
; FILING DATE: 12-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rohm, Benita J.  
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 313-496-7622  
; TELEFAX: 313-496-8454  
; INFORMATION FOR SEQ ID NO: 16:

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
; US-08-303-025-16

Query Match 35.0%; Score 74.5; DB 1; Length 33;
Best Local Similarity 68.8%; Pred. No. 0.039;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Qy 5 AKKAKEKAKYKAAEAKKAAKYEKAAAEKA 36
   ||||| :||||| :||||| :|||||
Db 3 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 31

RESULT 15
US-08-436-703B-4
; Sequence 4, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J. Rohm, Esq.
; STREET: 6601 Woodward Avenue
; STREET: Suite 1525
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk 1.44Mb, 3.5"
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 6;
; SOFTWARE: ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,703B
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: N/A
; FILING DATE: N/A
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: 7WK-060548-00233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-965-1976
; TELEFAX: 313-965-1951
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
```

```
;
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
; US-08-436-703B-4

Query Match 35.0%; Score 74.5; DB 2; Length 33;
Best Local Similarity 68.8%; Pred. No. 0.039;
Matches 22; Conservative 3; Mismatches 4; Indels 3; Gaps 2;

Qy 5 AKKAKEKAKYKAAEAKKAAKYEKAAAEKA 36
   ||||| :||||| :||||| :|||||
Db 3 AKKA-AKKAKKA--AKKAKKAAKKAKKAAKKA 31

Search completed: March 10, 2003, 12:30:02
Job time : 8.97468 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 16.4219 Seconds  
(without alignments)  
327.825 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKKAEKAAK.....EAKYKAEAKAAKAEAYEA 56  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	44.0	421	2 JV0057	tolA protein - Esc
2	116.5	43.8	239	2 AE1317	hypothetical prote
3	113	42.5	394	2 F90725	membrane spanning
4	113	42.5	394	2 G85576	membrane spanning
5	110.5	41.5	376	2 AG0592	tolA protein [impo
6	110	41.4	1701	2 T09127	probable erythrocy
7	105.5	39.7	372	2 G64064	outer membrane int
8	102.5	38.5	210	2 A25550	histone H1 - sea u
9	102.5	38.5	347	2 E83525	histone H1 - sea u
10	102	38.3	211	2 A28100	tolA protein PA097
11	100.5	37.8	1390	2 S51364	histone H1-beta, e
12	99.5	37.4	206	1 HSTR1R	sperm tail-specific
13	98.5	37.0	243	2 AE1699	histone H1 - rainb
14	98.5	37.0	311	2 T17698	hypothetical prote
15	97.5	36.7	206	2 S09388	hypothetical prote
16	97	36.5	219	2 B60110	histone H1 - sea u
17	96	36.1	384	2 B43592	repetitive protein
18	95.5	35.9	217	2 A26721	outer membrane pro
19	95.5	35.9	248	1 HSUR1P	histone H1-gamma,
20	95.5	35.9	1403	2 T11583	probable translati
21	95	35.7	388	2 AC0138	tolA colicin impor
22	94.5	35.5	214	2 G70673	probable hupB - My
23	94.5	35.5	328	2 A44993	cytosolic repetiti
24	94	35.3	433	2 S25194	zucotin - yeast (Sa
25	94	35.3	703	2 T48600	kinase-like protei
26	92.5	34.8	182	2 S61926	histone H1 homolog
27	92.5	34.8	1128	2 T30296	R27-2 protein - Tr
28	92	34.6	220	2 A28456	histone H1.10 - ch
29	92	34.6	236	2 S22322	histone H1 - wheat

histone H1-delta -  
histone H1 (clone  
histone H1-II - Vo  
conserved hypochet  
Iga-specific metal  
histone H1 - Chlam  
immediate-early pr  
translation initia  
probable late embr  
histone H1 - trout  
tolA protein VC183  
arylesterase-relat  
hypothetical prote  
hypothetical prote  
histone H1 (clone  
hypothetical prote

ALIGNMENTS

RESULT 1

JV0057

tolA protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 01-Mar-2002

C;Accession: JV0057; B64810

R;Levengood, S.K.; Webster, R.E.

J. Bacteriol. 171, 6600-6609, 1989

A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their products

A;Reference number: JV0057; MUID:90078104; PMID:2687247

A;Accession: JV0057

A;Molecule type: DNA

A;Residues: 1-421 <LEV>

A;Cross-references: GB:M28232; NID:gl48018; PIDN:AAA24683.1; PID:gl48019

A;Experimental source: strain JM105

A;Note: the authors translated the initiation codon GTG for residue 1 as Val

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64810

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-421 <BLAT>

A;Cross-references: GB:AE000177; GB:U00096; NID:gl786955; PIDN:AAC73833.1; PID:gl786960;

A;Experimental source: strain K-12, substrain MGI655

C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach the

C;Genetics:

A;Gene: tolA

A;Map position: 17 min

A;Start codon: GTG

C;Keywords: nucleotide binding; P-loop; transmembrane protein

F;14-34/Domain: transmembrane #status predicted <MSS>

F;78-301/Domain: helical #status predicted <HSR>

F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 44.0%; Score 117; DB 2; Length 421;

Best Local Similarity 54.7%; Pred. No. 0.0043;

Matches 35; Conservative 5; Mismatches 10; Indels 14; Gaps 2;

QY 7 KEKAYAKKAEAKKAEAKKAYK-----AAEAKKAEAA-----KYKAEAAKAAKAEAA 52

Db 148 KAEADAKAEAAKAAADAKKAE 207

QY 53 AYEAA 56

Db 208 RKKA 211

RESULT 2

AE1317



**Qy**

2 KKYAKKEKAYAKKAEEAAKKAEAKYAAEAK-----KKAEAKYKA 42  
| : | : ||||: |||| | |||:

**Dd**

96 KTTRAQARAARAAKKLAARKKEQEKEKKAATKTARKEKLAAKCAAKAACKVKKPAPPAKAKK 155

sperm tail-specific protein mst101(2) - fruit fly (*Drosophila hydei*)





THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 12.7595 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 56

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	44.0	421	1	TOLA_ECOLI
2	105.5	39.7	372	1	TOLA_HAEIN
3	105	39.5	349	1	RS6_AEDAL
4	104.5	39.3	346	1	RS6_AEDAE
5	102.5	38.5	210	1	H1_LYTP1
6	102.5	38.5	347	1	TOLA_PSEAE
7	100.5	37.8	1391	1	MST2_DROHY
8	99.5	37.4	206	1	H1_OCMY
9	99.5	37.4	237	1	H1E_CHITE
10	99	37.2	211	1	H1B_STRPU
11	97.5	36.7	205	1	DBH_MYCBO
12	96	36.1	384	1	TMPE_TREPH
13	95.5	35.9	217	1	H1G_STRPU
14	95.5	35.9	248	1	H1_PARAN
15	95.5	35.9	1403	1	YDF3_SCHPO
16	95	35.7	917	1	IF2_PROVU
17	94.5	35.5	214	1	DBH_MYCTU
18	94	35.3	433	1	ZUO1_YEAST
19	93.5	35.2	208	1	DBH_MYCSM
20	92	34.6	219	1	H1O_CHICK
21	91.5	34.4	185	1	H1D_STRPU
22	91.5	34.4	238	1	H1_WHEAT
23	91	34.2	139	1	ASR_KLEPN
24	91	34.2	221	1	H1C_CHITE
25	91	34.2	240	1	H12_VOLCA
26	90.5	34.0	235	1	H1E_CHIPA
27	90.5	34.0	1532	1	IGA_NEIGO
28	90	33.8	407	1	IF68_HSVSA
29	90	33.8	884	1	IF2_YERPE
30	89.5	33.6	194	1	H1_SALTR
31	88.5	33.3	233	1	H1I_GLYSA
32	88	33.1	771	1	CALD_CHICK
33	87.5	32.9	233	1	H1I_GLYBA

RESULT 1  
TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levensgood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of their products, components of a multistep translocation system in Escherichia coli.";  
RT J. Bacteriol. 171:6600-6609(1989).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:12453-1474(1997).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
[4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "Tola: a membrane protein involved in colicin uptake contains an extended helical region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
[5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

## ALIGNMENTS

34 87 32.7 344 1 MST1\_DROHY Q08695 drosophila  
35 86.5 32.5 218 1 H1O1\_CHICK P08284 gallus gall  
36 86 32.3 218 1 DBH2\_STRCO O86537 streptomyc  
37 85.5 32.1 232 1 H1B\_CHITE P40276 chironomus  
38 85.5 32.1 325 1 TMPE\_TREPA P19649 treponema p  
39 85 32.0 135 1 HCT1\_CHLTR C02281 chlamydia t  
40 85 32.0 171 1 H1E\_CHCR P02257 echinolampa  
41 85 32.0 217 1 H1\_CHICK P09937 gallus gall  
42 85 32.0 244 1 H1O\_CHITH Q07134 chironomus  
43 84.5 31.8 952 1 TOPI\_STRCO Q9X909 streptomyc  
44 84 31.6 184 1 RS16\_BACTN Q9RQ15 bacteroides  
45 84 31.6 207 1 H1I\_CABEL P10771 caenorhabdi

RA Llobes R.;  
RT "TOLA central domain interacts with Escherichia coli porins.";  
RL EMO J. 15:6408-6415(1996).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.  
RX MEDLINE=99332679; PubMed=10404600;  
RA Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;  
RT "Filamentous phage infection: crystal structure of g3p in complex  
with its coreceptor, the C-terminal domain of Tola.";  
RL Structure 7:711-722(1999).  
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A  
COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE  
BINDING TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION  
OF BACTERIOPHAGE DNA.  
CC -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE  
AND LAMB.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M28232; AAA24683.1; -;  
DR EMBL; AE000177; AAC73833.1; -;  
DR EMBL; D90713; BAA35405.1; -;  
DR PIR; JVO057; JVO057.  
DR PDB; 1TOL; 20-MAY-99.  
DR EcoGene; EG11007; tola.  
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;  
Repeat; Inner membrane; 3D-structure; Complete proteome.  
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 14 34 POTENTIAL.  
FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).  
FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).  
FT DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2) -  
A(2,4).  
FT SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;  
Query Match 44.0%; Score 117; DB 1; Length 421;  
Best Local Similarity 54.7%; Pred. No. 0.0024;  
Matches 35; Conservative 5; Mismatches 10; Indels 14; Gaps 2;  
Qy 7 KEKAYAKAEAKKAEAKAYK-----AAEAKKAE-----KYKAEAKAAKAE 52  
Db 148 KAEADAKAEAEAKKAAADAKKAEAEAKAEAKAEAKAEAKAEAKAEAKAEAKAE 207  
Qy 53 AYE 56  
Db 208 RKKA 211  
RESULT 2  
TOLA HAEN STANDARD; PRT; 372 AA.  
AC P44678; P94810;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE TOLA protein.  
GN TOLA OR H10383.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN= / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kervlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.W.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1479;  
RX MEDLINE=97080550; PubMed=8921895;  
RA Sen K., Sikkema D.J., Murphy T.F.;  
RT "Isolation and characterization of the Haemophilus influenzae toIQ,  
RT toIR, toLA and toLB genes.";  
RL Gene 178:75-81(1996).  
CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS  
(BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
(Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U32722; AAC22041.1; -;  
DR EMBL; U32470; AAC44596.1; -;  
DR HSP; P19934; 1TOL.  
DR TIGR; H10383; -;  
KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;  
Complete proteome.  
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 9 29 POTENTIAL.  
FT DOMAIN 30 372 PERIPLASMIC (POTENTIAL).  
FT VARIANT V -> A (IN STRAIN 1479).  
FT VARIANT K -> R (IN STRAIN 1479).  
FT VARIANT A -> P (IN STRAIN 1479).  
FT VARIANT A -> R (IN STRAIN 1479).  
FT VARIANT V -> A (IN STRAIN 1479).  
FT VARIANT D -> A (IN STRAIN 1479).  
FT VARIANT T -> A (IN STRAIN 1479).  
FT VARIANT T -> A (IN STRAIN 1479).  
FT VARIANT L -> F (IN STRAIN 1479).  
FT VARIANT I -> V (IN STRAIN 1479).  
FT VARIANT N -> S (IN STRAIN 1479).  
FT VARIANT T -> A (IN STRAIN 1479).  
FT VARIANT S -> P (IN STRAIN 1479).  
FT SQ SEQUENCE 372 AA; 39831 MW; 266ECF05C6C95544 CRC64;  
Query Match 39.7%; Score 105.5; DB 1; Length 372;  
Best Local Similarity 48.2%; Pred. No. 0.019;  
Matches 27; Conservative 14; Mismatches 14; Indels 1; Gaps 1;  
Qy 2 KYAKKAEKAYAKAEKAAK-KAEAKYKAAAEAKKAEKAYKAEAKAAKAEKAAE 56  
Db 133 EKQQAEEAKAKQAEEAKAKLKADEAKRLAAAKQAEEAKAEKAAEIAAKQAK 188  
RESULT 3  
RS6\_AEDAL  
ID RS6\_AEDAL STANDARD; PRT; 349 AA.  
AC Q90762;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)









FT	DOMAIN	1	90	BACTERIAL HISTONE-LIKE DOMAIN.
FT	DOMAIN	101	205	DEGENERATE REPEATS REGION.
FT	CONFLICT	199	199	A -> T (IN REF. 2).
SQ	SEQUENCE	205 AA;	21262 MW;	19FC67885DFB6A8 CRC64;
 Query Match 36.7%; Score 97.5; DB 1; Length 205; Best Local Similarity 56.9%; Pred.No. 0.055; Matches 29; Conservative 4; Mismatches 15; Indels 3; Gaps				
QY	1 AKVYAKKAYAKKAQAQAAKAYKAAEAKKKABAKYKAEAKKAAKAAKE 151   :    Db 111 AKVKAK--KAPAKATYAKAAKATKA-PAKAAATKAPAKAVKATKSPAKK 158   :			
 RESULT 12				
ID	TMPB_TREPH	STANDARD;	PRT;	384 AA.
AC	E297Z0;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Treponemal membrane protein B precursor (Antigen tmpB).			
GN	TMPB.			
OS	Treponema phagedenis.			
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.			
ON	NCBI_TaxID=162;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=Kazan 5;			
RC	MEDLINE=91372963; PubMed=1894368;			
RA	Yelton D.B.; Limberger R.J.; Curci K., Malinosky-Rummell F.,			
RA	Silivensky L., Schouls L.M., van Embden J.D., Charon N.W.;			
RT	"Treponema phagedenis encodes and expresses homologs of the Treponema			
RT	pallidum TmpA and TmpB proteins.";			
RL	Infect. Immun. 59:3685-3693(1991).			
CC	-!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR			
CC	LARGE MOLECULES.			
CC	-!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.			
CC	-!- SIMILARITY: TO TMPB OF T.PALLIDUM.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collabor-			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstat-			
CC	the European Bioinformatics Institute. There are no restrictions on co-			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commo-			
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M58563; AAA27480.1; -.			
DR	PIR; B43592; B43592.			
KW	Antigen; Outer membrane; Repeat; Signal.			
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	384	TREPONEMAL MEMBRANE PROTEIN B.
FT	DOMAIN	151	235	17 X 5 AA TANDEM REPEATS OF K-A-A-[AKR]-[ED].
FT	REPEAT	151	155	1-1.
FT	REPEAT	156	160	1-2.
FT	REPEAT	161	165	1-3.
FT	REPEAT	166	170	1-4.
FT	REPEAT	171	175	1-5.
FT	REPEAT	176	180	1-6.
FT	REPEAT	181	185	1-7.
FT	REPEAT	186	190	1-8.
FT	REPEAT	191	195	1-9.
FT	REPEAT	196	200	1-10.
FT	REPEAT	201	205	1-11.
FT	REPEAT	206	210	1-12.
FT	REPEAT	211	215	1-13.
FT	REPEAT	216	220	1-14.
FT	REPEAT	221	225	1-15.
FT	REPEAT	226	230	1-16.
FT	REPEAT	231	235	1-17.
FT	DOMAIN	236	288	6 X 8 AA TANDEM REPEATS OF [EA]-A-A-R-X-

```

FT REPEAT      236      243      A-A-E.
FT REPEAT      245      252      2-1.
FT REPEAT      254      261      2-2.
FT REPEAT      263      270      2-3.
FT REPEAT      272      279      2-4.
FT REPEAT      281      288      2-5.
FT REPEAT      284      291      2-6.
SQ SEQUENCE    384 AA; 42677 MW; 6B94CBC74294DE8C CRC64;

Query Match      36.1%; Score 96; DB 1; Length 384;
Best Local Similarity 51.6%; Pred. No. 0.12;
Matches 33; Conservative 6; Mismatches 17; Indels 8; Gaps 3;

Qy 1 AKKYAKKEKAYAKKA--EKAARK--AEAKAYKAAEAKKA-----EAKYKAAEAKAAKEA 52
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 AKEKAKEKAADKAKEKAKEKAADKAKEKAKEKAKEKAKEKAKEKAKEKAKEKAKEKA 237
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 53 AYE 56
    ||| |||
Db 238 ARKA 241

RESULT 13
HIG_STRPU
ID_HIG_STRPU STANDARD; PRT; 217 AA.
AC P07796;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-gamma, late.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87172742; PubMed=3031476;
RA Knowles J.A., Lai Z.-C., Childs G.J.;
RT "Isolation, characterization, and expression of the gene encoding the
RT late histone subtype H1-gamma of the sea urchin Strongylocentrotus
RT purpuratus."
RL Mol. Cell. Biol. 7:478-485 (1987).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M16033; AAA30059.1; -.
DR PIR; A26721; A26721.
DR HSP; P02259; IHST.
DR InterPro; IPR001386; Histone H1/H5.
DR Pfam; PF00538; Linker histone N.
DR ProDom; PD000373; Linker histone; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE    217 AA; 22658 MW; C7251EED3413B185 CRC64;

Query Match      35.9%; Score 95.5; DB 1; Length 217;
Best Local Similarity 43.7%; Pred. No. 0.084;
Matches 31; Conservative 5; Mismatches 16; Indels 19; Gaps 2;

Qy 1 AKKYAKK-----EKAAYAKAEK-----AAKAEAKAYKAAEAKKAEAKYK 41
    ||| ||| : ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 135 AKKAATKATKTKKVPAAKKAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 194
Qy 42 AEAKAAKAAKEA 52
    ||| ||| ||| |||
Db 195 KKAAPAKKAA 205

RESULT 14
H1_PARAN
ID_H1_PARAN STANDARD; PRT; 248 AA.
AC P02256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
OS Parachinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinidae;
OC Parachinus.
OX NCBI_TaxID=7658;
RN [1]
RP SEQUENCE OF 1-84.
RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanoogen bromide peptides."
RL Eur. J. Biochem. 104:559-566 (1980).
RN [2]
RP SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure."
RL Eur. J. Biochem. 104:567-578 (1980).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC PIR; A02586; HSURJP.
DR HSP; P02259; IHST.
DR InterPro; IPR001386; Histone H1/H5.
DR Pfam; PF00538; Linker histone; 1.
DR ProDom; PD000373; Linker histone; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Sperm.
FT VARIANT      144      144      K -> R.
SQ SEQUENCE    248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match      35.9%; Score 95.5; DB 1; Length 248;
Best Local Similarity 52.9%; Pred. No. 0.093;
Matches 27; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

Qy 1 AKKYAKKEKAYAKKAKEKA--KAEAKAYKAAEAKKAEAKYKAEAKYKAEAKYKAEAK 50
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 AKAAKKAALAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 195
    ||| ||||| ||| ||||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
YDF3_SCHPO
ID_YDF3_SCHPO STANDARD; PRT; 1403 AA.
AC Q10475;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable eukaryotic initiation factor C17C9.03.

```

Search completed: March 10, 2003, 12:17:09  
Job time: 13.7595 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 29.654 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-3

Perfect score: 266

Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYAEAAKAAKEAAEAYEA 56

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTREMBL 21.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_mhc.\*

9: sp\_organelle.\*

10: sp\_phase.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp\_vertebrate.\*

15: sp\_unclassified.\*

16: sp\_rvirus.\*

17: sp\_bacteriap.\*

18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	117	44.0	372	Q9WXX1	Q9wxx1 pseudomonas
2	117	44.0	389	16 Q9CM70	Q9cm70 pasteurella
3	116.5	43.8	239	16 Q8YSW4	Q8ysw4 listeria mo
4	114.5	43.0	395	2 Q937K4	Q937k4 erwinia chr
5	113.5	42.7	407	16 Q8ZQT6	Q8zqt6 salmonella
6	113	42.5	394	16 Q8X965	Q8x965 escherichia
7	111	41.7	212	3 Q91946	Q91946 candida alb
8	110.5	41.5	376	16 Q8Z8C1	Q8z8c1 salmonella
9	110	41.4	1701	5 Q61164	Q61164 plasmodium
10	104.5	39.3	1866	5 Q8T5C8	Q8t5c8 plasmodium
11	100.5	37.8	200	16 Q8XVW7	Q8xvw7 ralstonia s
12	98.5	37.0	243	16 Q92A67	Q92a67 listeria in
13	98.5	37.0	311	12 Q84528	Q84528 paramecium
14	98	36.8	969	5 Q9ND19	Q9nd19 plasmodium
15	96	36.1	213	3 Q9UV33	Q9uv33 ascobolus i
16	95.5	35.9	629	2 Q9KJ98	Q9kj98 escherichia

17	95	35.7	275	10 Q9XHL9	Q9xhl9 triticum ae
18	95	35.7	388	16 Q8ZGZ2	Q8zgz2 yersinia pe
19	94.5	35.5	845	5 Q9YIP8	Q9yip8 plasmodium
20	94	35.3	191	5 Q46141	Q46141 mytilus edu
21	94	35.3	191	5 Q46142	Q46142 mytilus edu
22	94	35.3	191	5 Q46143	Q46143 mytilus edu
23	94	35.3	191	5 Q46362	Q46362 mytilus edu
24	94	35.3	703	10 Q9LYA2	Q9lya2 arabidopsis
25	94	35.3	1341	16 Q98KG7	Q98kg7 rhizobium l
26	93	35.0	275	5 Q01395	Q01395 drosophila
27	93	35.0	688	10 Q93VS6	Q93vs6 arabidopsis
28	92.5	34.8	182	2 Q45370	Q45370 bordetella
29	92.5	34.8	1128	5 Q26947	Q26947 trypanosoma
30	92.5	34.8	1671	5 Q8T5C9	Q8t5c9 plasmodium
31	92	34.6	233	5 Q18319	Q18319 chironomus
32	92	34.6	243	2 Q9L564	Q9l564 streptococc
33	92	34.6	247	2 Q9L566	Q9l566 streptococc
34	92	34.6	401	2 Q9LAZ2	Q9laz2 streptococc
35	92	34.6	2055	5 Q8T5C7	Q8t5c7 plasmodium
36	91.5	34.4	236	10 Q9SMU3	Q9smu3 triticum ae
37	91.5	34.4	237	10 Q9SMU2	Q9smu2 triticum ae
38	91.5	34.4	284	10 Q65794	Q65794 triticum ae
39	91	34.2	191	5 Q46140	Q46140 mytilus edu
40	90.5	34.0	581	16 Q9EU45	Q9eu45 deinococcus
41	90.5	34.0	755	10 Q9FP71	Q9fp71 oryza sativ
42	90	33.8	101	5 Q9BMY8	Q9bmy8 leishmania
43	90	33.8	345	16 Q8Y1F6	Q8y1f6 ralstonia s
44	89.5	33.6	136	10 Q39681	Q39681 daucus caro
45	89.5	33.6	356	16 Q9KX10	Q9kx10 vibrio chol

## ALIGNMENTS

### RESULT 1

Q9WXX1	ID	Q9WXX1	PRELIMINARY;	PRT;	372 AA.
AC	Q9WXX1;				
DT	01-NOV-1999	(TrEMBLrel. 12, Created)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)			
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)			
DE	Tola protein.				
GN	TOLA.				
OS	Pseudomonas putida.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;				
OC	Pseudomonas.				
OX	NCBI_TaxID=303;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MT-2;				
RX	MEDLINE=96198174; PubMed=8626299;				
RA	Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.J.				
RT	"The Pseudomonas putida peptidoglycan-associated outer membrane				
RT	lipoprotein (PAL) is involved in maintenance of the integrity of the				
RT	cell envelope."				
RL	J. Bacteriol. 178:1699-1706(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MT-2;				
RA	Ramos-Gonzalez I.;				
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MT-2;				
RA	Rodriguez-Herva J.J.;				
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MT-2;				
RX	MEDLINE=96422022; PubMed=8824639;				
RA	Rodriguez-Herva J.J., Ramos J.J.;				
RT	"Characterization of an OprL null mutant of Pseudomonas putida."				
RL	J. Bacteriol. 178:5836-5840(1996).				

RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordstrek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
RA Varquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;  
RT "Comparative genomics of *Listeria* species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591981; CAD00019.1; -.  
DR Listalist; LMO01941; -.  
DR InterPro; IPR002482; LysM.  
DR Pfam; PF01476; LysM; 1.  
DR SMART; SM00257; LysM; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 239 AA; 25836 MW; 72E59D576E0D7832 CRC64;

Query Match 43.8%; Score 116.5; DB 16; Length 239;  
Best Local Similarity 55.4%; Pred. No. 0.0026;  
Matches 31; Conservative 9; Mismatches 15; Indels 1; Gaps 1;

OY 1 AKKYAKKEKAYAKAEAAKKAAEKAAEKAAEKAAEKAAEKAAEKAAEAEE 55  
Db 116 AEKAAEEKAAEAADKKQEDAVKANAARKEQAEEKAADKAAAEKAAEA 171

RESULT 4  
OY37K4 PRELIMINARY; PRT; 395 AA.  
AC OY37K4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE TOLA protein.  
GN TOLA.  
OS Erwinia chrysanthemi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Pectobacterium.  
OX NCBI\_TaxID=556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3937;  
RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;  
RT "Characterization of the Erwinia chrysanthemi tol-pal genes.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ297885; CAC82708.1; -.  
SQ SEQUENCE 395 AA; 41601 MW; 3C0C1DC12E181013 CRC64;

Query Match 43.0%; Score 114.5; DB 2; Length 395;  
Best Local Similarity 54.2%; Pred. No. 0.0063;  
Matches 32; Conservative 11; Mismatches 13; Indels 3; Gaps 2;

OY 1 AKKYAKKE-KAYAKAEAAKKA--EAKYKAAEAACKKAAEKYKAAEAACKKAAEAEEA 56  
Db 182 AKKAAEEAAKTAAAAAEAKKAAEEAAKAAADAKQAAEEAAKAAEAACKKAAEAADA 240

RESULT 5  
OY37Q6 PRELIMINARY; PRT; 407 AA.  
AC OY37Q6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Tol protein, membrane spanning protein.  
GN TOLA OR STM0747.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,



```
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
DR EMBL: AE008730; AAL19691.1; -.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;

Query Match 42.7%; Score 113.5; DB 16; Length 407;
Best Local Similarity 49.3%; Pred. No. 0.0079;
Matches 37; Conservative 10; Mismatches 9; Indels 19; Gaps 4;

Qy 1 AKKAKKAKKAKKAKKAA---KKAERAKYK-AAEAKKKAAE-----KYKAAEA 44
Db 123 AAKLAQQQQQAEAAKAAADAAKKAERAKKAAEAAKAAADAAKKAERAKKAAEAA 182

Qy 45 AKAA---KEAAEYA 56
Db 183 AKAAADAKKAAEAA 197

RESULT 6
Q8X965 PRELIMINARY; PRT; 394 AA.
AC Q8X965;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Membrane spanning protein, required for outer membrane integrity
DE (Membrane spanning protein TolA).
DE TOLA OR Z0907 OR ECS0774.
GN Escherichia coli O157:H7.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005252; AAG55075.1; -.
DR EMBL: AP002553; BAB34197.1; -.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5858D8E8230BDE28 CRC64;

Query Match 42.5%; Score 113; DB 16; Length 394;
Best Local Similarity 60.0%; Pred. No. 0.0085;
Matches 33; Conservative 5; Mismatches 11; Indels 6; Gaps 2;
```

```
Qy 8 EKAYAKKAAKAAKAAKAYK-AAEAKKKAAE-----KYKAAEAAKAAKAAEAYA 56
Db 157 EEAAGAADAADAKKAAEAAKAAEAAKAAEAAKAAEAAKAAEAAEAAEARKKA 211

RESULT 7
Q93946 PRELIMINARY; PRT; 212 AA.
AC Q93946;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CTA2p (Fragment).
GN CTA2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaiser B., Kunkel W., Saluz H.P., Munder T.;
RT "Identification of Candida albicans protein domains with
RT transcriptional activating properties in Saccharomyces cerevisiae.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ006637; CAA07165.1; -.
FT NON TER 1
SQ SEQUENCE 212 AA; 24231 MW; 10C2122E9554A387 CRC64;

Query Match 41.7%; Score 111; DB 3; Length 212;
Best Local Similarity 55.9%; Pred. No. 0.0071;
Matches 33; Conservative 5; Mismatches 17; Indels 4; Gaps 2;

Qy 1 AKK---YAKKEKAVAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAEYA 56
Db 63 AKKKEEAKKAAEAKKKEEAKKAE-EAKKAAEAKKVEEAKKAAEAKKAEERAKKA 120

RESULT 8
Q8Z8C1 PRELIMINARY; PRT; 376 AA.
AC Q8Z8C1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tola protein.
GN STY0793.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jageis K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AL627268; CAD05209.1; -.
DR InterPro: IPR000104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
DR PRINTS: PR01574; TUBBYPROTEIN.
KW Complete proteome.
SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;
```



```

RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN EMBL; U42580; AAC96576.1; -
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
SQ SEQUENCE 311 AA; 35390 MW; 97B059B82AFEF88A CRC64;

Query Match 37.0%; Score 98.5; DB 12; Length 311;
Best Local Similarity 43.9%; Pred. No. 0.13;
Matches 29; Conservative 8; Mismatches 18; Indels 11; Gaps 2;

QY 2 KKYAKKEKAYAK-----KAEKAACKAEKAYKAAEAKK-----KAEKAYKAAEAKAAK 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 72 KERVYAKAAAREAKERVYKAAEAKERIKAEKAVKERIKAEKAAEKERVYKAEKAPAR 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 51 EAAEYEA 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 132 EAAKAA 137

RESULT 14
Q9NDI9 Q9NDI9 PRELIMINARY; PRT; 969 AA.
AC Q9NDI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 3g.
GN MSP-3g.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BELEM;
RX MEDLINE=21273153; PubMed=11377738;
RA Galinski M.R., Ingravallo P., Corredor-Medina C., Al-Khedery B.,
RV Povaia M., Barnwell J.W.;
RT "Plasmodium vivax merozoite surface proteins-3beta and-3gamma share
RT structural similarities with P. vivax merozoite surface protein-3alpha
RT and define a new gene family.";
RL Mol. Biochem. Parasitol. 115:41-53 (2001).
DR EMBL; AF099663; AAF78288.1; -.
SQ SEQUENCE 969 AA; 104674 MW; 9A150CCA9918FF77 CRC64;

Query Match 36.8%; Score 98; DB 5; Length 969;
Best Local Similarity 39.4%; Pred. No. 0.44;
Matches 28; Conservative 11; Mismatches 17; Indels 12; Gaps 1;

QY 1 AKKYAKKEKAYAKAEKAAKAEKAYKAAEAKKKAEAK-----YKAEAAKAA 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 AKKQAEKAEKTTATTEANKAKEEAKAEKAEKAEKAEKAEKAGDVDEYVAVNVFESVKAA 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 49 AKENAY 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 AKAAAH 659

RESULT 15
Q9UV33 Q9UV33 PRELIMINARY; PRT; 213 AA.
AC Q9UV33;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Histone H1.
GN H1.
OS Ascolobolus immersus.

```

```

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascobolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM42;
RX MEDLINE=20063166; PubMed=10594009;
RA Barra J.L., Rhounim L., Roessignol J.L., Faugeron G.;
RT "Histone H1 is dispensable for methylation-associated gene silencing
in Ascobolus immersus and essential for long life span.";
RL Mol. Cell. Biol. 20:61-69(2000).
DR EMBL; AF190622; AAF16011.1; -.
DR HSSP; P02259; IHST.
DR InterPro; IPR01386; Histone_H1/H5.
DR InterPro; IPR03216; Linkerhist_N.
DR Pfam; PF00538; linker_histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR PRODOM; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
SQ SEQUENCE 213 AA; 21897 MW; 3760C57536947BF8 CRC64;

Query Match      36.1%; Score 96; DB 3; Length 213;
Best Local Similarity 44.8%; Pred. No. 0.15;
Matches 30; Conservative 4; Mismatches 17; Indels 16; Gaps 1;

Qy 3 KYAKKEKAYAKKAEEAKKA-----EAKAYKAAEAKKAEAKYKAEAAK 46
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 94 KLAKKEKAAAPKPKPAKKAAPKDAAPKKAAPKKAAPKKAAPKSAAKKLLIDAKKAAAK 153
   | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 47 AAAYEAA 53
   | | | | |
Db 154 PAAKKA 160

```

Search completed: March 10, 2003, 12:25:55  
 Job time : 30.654 secs

Result No.	Score	Query Match	Length	ID	Description
1	266	100.0	56	21 AAY82573	Copolymer molecule
2	245.5	92.3	77	21 AAY82575	Copolymer molecule
3	233	87.6	86	21 AAY82576	Copolymer molecule
4	198	74.4	66	21 AAY82574	Copolymer molecule
5	180.5	67.9	109	21 AAY82577	Copolymer molecule
6	134.5	50.6	45	21 AAY82572	Copolymer molecule
7	116.5	43.8	239	23 ABA49123	Listeria monocytog
8	111	41.7	106	11 AAR06446	Recombinant copoly
9	104.5	39.3	35	21 AAY82571	Copolymer molecule
10	104	39.1	100	21 AAY38499	Peptide #10 used i

(TEVA-) TEVA PHARM USA INC.





AA82577  
 ID AAY82577 standard; peptide; 109 AA.  
 XX  
 AC AAY82577;  
 XX  
 DT 28-JUL-2000 (first entry)  
 XX  
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ\_ID NO:7.  
 XX  
 KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200018794-A1.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 24-SEP-1999; 99WO-US22402.  
 XX  
 PR 25-SEP-1998; 98US-0101693.  
 XX  
 PA (VEDA ) YEDA RES & DEV CO LTD.  
 PA (TEVA-) TEVA PHARM USA INC.  
 XX  
 PI Gad A, Lis D;  
 XX  
 DR WPI; 2000-317499/27.  
 XX  
 PT Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glatiramer acetate and for treatment and prevention of immune diseases  
 XX  
 PS Claim 10; Page 14; 72pp; English.  
 XX  
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions,  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights and physical properties which are analogous to  
 CC glatiramer acetate markers.  
 CC molecular weight markers.  
 XX  
 SQ Sequence 109 AA;  
 Query Match 67.9%; Score 180.5; DB 21; Length 109;  
 Best Local Similarity 45.9%; Pred. No. 2.6e-11;  
 Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3;  
 QY 1 AKYAKK-EKAYAKA-----EKAAKAEKAYKAAEKKAEAK----- 39

Db 1 AKYAKAEKAYAKAEKAYAKAEKAYKAAEKKAEKAYKAAEKKAEKAYKAAEKKAEK 60  
 QY 40 -----YKAEAAKAAKAAEAA 56  
 Db 61 YKAEAKKAYAKAEKAYKAAEKKAEKAYKAAEKKAEKAYKAAEKKAEKAAEAA 109  
 RESULT 6  
 AAY82572  
 ID AAY82572 standard; peptide; 45 AA.  
 XX  
 AC AAY82572;  
 XX  
 DT 28-JUL-2000 (first entry)  
 XX  
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ\_ID NO:2.  
 XX  
 KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200018794-A1.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 24-SEP-1999; 99WO-US22402.  
 XX  
 PR 25-SEP-1998; 98US-0101693.  
 XX  
 PA (VEDA ) YEDA RES & DEV CO LTD.  
 PA (TEVA-) TEVA PHARM USA INC.  
 XX  
 PI Gad A, Lis D;  
 XX  
 DR WPI; 2000-317499/27.  
 XX  
 PT Copolymer 1 related polypeptides used as molecular weight markers for  
 PT glatiramer acetate and for treatment and prevention of immune diseases  
 XX  
 PS Claim 10; Page 14; 72pp; English.  
 XX  
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions,  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights and physical properties which are analogous to  
 CC glatiramer acetate markers.  
 CC molecular weight markers.  
 XX





AAV92571 to AAV09577 represent specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glutarimer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, sclerosing, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated diseases which can be treated

```
Query Match          39.1%; Score 104; DB 21; Length 100;
Best Local Similarity 58.2%; Pred.No. 0.0013;
Matches 32; Conservative 4; Mismatches 17; Indels 2; Gaps 2
```

QY 1 AKTYAK-KEKAYAKAKEAAKKAKAKAYAKAEAKKKBAKYKABA-AKAAAKEAA 53  
Db 2 AKA 56

RESULT 12  
AAU04289  
ID AAU04289 standard; Peptide; 100 AA.  
XX  
XX AC  
XX AAU04289;  
XX  
DT 23-OCT-2001 (first entry)

xx	
DE	Poly-Lys-Ala used in nucleic acid transporter system.
xx	
KW	Nucleic acid transport; cytosol; ligand; lysis agent; spacer molecule;
KW	gene therapy; hepatocyte; muscle; bone forming cell.
xx	
xx	

OS	
XX	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 3..100
FT	/note= "Iys-Ala in positions 3-100 may be present or absent"
FT	
XX	

FN	US6177534-B1.
XX	
PD	23-JAN-2001.
XX	
XX	
PF	05-JUN-1995; 95US-0462040.
XX	
XX	
PR	14-DEC-1993; 93US-0167641.
PR	20-MAR-1992; 92US-0855389.
PR	19-MAR-1993; 93WO-US02725.
XX	
XX	
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.

PI Woo SLC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;  
XX  
XX WPI; 2001-365933/38.  
XX  
XX Nucleic acid transport system. useful for creating transgenic animals

XX  
PS  
XX

Disclosure; Column 131; 111pp; English.

XX  
CC

The sequence represents poly-Ivs-Ala. used to bind nucleic acid in a

CC nucleic acid transporter system. The nucleic acid transporter system uses  
CC nucleic acid binding complexes containing surface ligands which are  
CC capable of binding to a cell surface receptor and entering the cell  
CC through cytosol. The compounds of the invention are either ligands,  
CC binding molecules (surface ligands), lysis agents, spacer molecules or  
CC their intermediates. The ligands, binding molecules, lysis agents and  
CC spacer molecules are used in nucleic acid transporter systems to deliver  
CC nucleic acid into specific cells e.g. in gene therapy to deliver nucleic  
CC acid into hepatocytes, muscle cells or bone forming cells.





**THIS PAGE BLANK (USPTO)**













THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 9.92405 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-3  
Perfect score: 266  
Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAAKAAKAEAYEA 56

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pdp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pdp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pdp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pdp.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pdp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	39.1	56	4	US-08-993-008A-6
2	104	39.1	100	2	US-08-460-890A-64
3	104	39.1	100	3	US-08-167-641C-64
4	104	39.1	100	4	US-08-460-371A-64
5	104	39.1	100	4	US-08-462-040-64
6	103	38.7	223	4	US-09-095-855-201
7	103	38.7	223	4	US-09-205-426-201
8	101	38.0	1507	3	US-08-929-329-5
9	99.5	37.4	48	4	US-08-993-008A-5
10	94.5	35.5	214	3	US-09-041-889-27
11	94	35.3	433	1	US-08-346-849-2
12	94	35.3	433	2	US-08-293-284A-2
13	92.5	34.8	643	2	US-08-216-894-8
14	92.5	34.8	643	4	US-09-115-746-8
15	91.5	34.4	60	1	US-08-346-849-16
16	91.5	34.4	60	2	US-08-293-284A-16
17	91.5	34.4	472	2	US-08-216-894-10
18	91.5	34.4	472	4	US-09-115-746-10
19	91.5	34.4	564	2	US-08-216-894-2
20	91.5	34.4	564	4	US-09-115-746-2
21	90.5	34.0	1507	6	5268270-2
22	84	31.6	32	1	US-08-152-488-13
23	84	31.6	32	1	US-08-303-025-15
24	84	31.6	32	1	US-08-677-304-13
25	84	31.6	32	2	US-08-436-703B-2
26	84	31.6	33	1	US-08-303-025-16
27	84	31.6	33	2	US-08-436-703B-4

28 82 30.8 288 3 US-08-312-949-4 Sequence 4, Appli  
29 82 30.8 288 3 US-08-446-201-4 Sequence 4, Appli  
30 82 30.8 289 1 US-08-072-070-4 Sequence 4, Appli  
31 82 30.8 289 1 US-08-469-434-4 Sequence 4, Appli  
32 82 30.8 289 1 US-08-214-222-4 Sequence 4, Appli  
33 82 30.8 289 2 US-08-467-852A-5 Sequence 5, Appli  
34 82 30.8 289 2 US-08-468-718-4 Sequence 4, Appli  
35 82 30.8 289 2 US-08-247-491A-5 Sequence 5, Appli  
36 82 30.8 619 1 US-08-465-746-2 Sequence 2, Appli  
37 82 30.8 619 1 US-08-214-164-2 Sequence 2, Appli  
38 82 30.8 619 2 US-08-467-852A-3 Sequence 3, Appli  
39 82 30.8 619 2 US-08-246-636-2 Sequence 2, Appli  
40 82 30.8 619 2 US-08-247-491A-3 Sequence 3, Appli  
41 82 30.8 619 2 US-08-319-795-2 Sequence 2, Appli  
42 82 30.8 619 2 US-08-468-985-2 Sequence 2, Appli  
43 82 30.8 619 3 US-08-312-949-2 Sequence 2, Appli  
44 82 30.8 648 1 US-08-072-070-2 Sequence 2, Appli  
45 82 30.8 648 1 US-08-469-434-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-993-008A-6  
; Sequence 6, Application US/08993008A  
; Patent No. 6153596  
; GENERAL INFORMATION:  
; APPLICANT: Liotta, Dennis C.  
; APPLICANT: Petros, John A.  
; APPLICANT: Wey, Shioh-Jyi  
; APPLICANT: Karr, Joan F.  
; APPLICANT: Pohl, Jan  
; TITLE OF INVENTION: Polycationic Oligomers  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/993,008A  
; FILING DATE: 18-DEC-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/032,436  
; FILING DATE: 18-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sullivan, Sally A.  
; REGISTRATION NUMBER: 32,064  
; REFERENCE/DOCKET NUMBER: 33-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303-499-8080  
; TELEFAX: 303-499-8089  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 56 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-993-008A-6

Query Match 39.1%; Score 104; DB 4; Length 56;



RESULT 5  
US-08-462-040-64  
; Sequence 64, Application US/08462040  
; Patent No. 6177554  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street

```

/
;
; MEDIUM TYPE: Diskette
;
; COMPUTER: IBM Compatible
;

```

```

/
;
; MEDIUM TYPE: Diskette
;
; COMPUTER: IBM Compatible
;

```









NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-8

Query Match 34.8%; Score 92.5; DB 4; Length 643;  
Best Local Similarity 47.5%; Pred. No. 0.039;  
Matches 29; Conservative 6; Mismatches 15; Indels 11; Gaps 2;

Qy 1 AKYAKKEKAYAKAEKAAKAEKAYKAEAKKAEAKYK-----AEAAKAAAEK 52  
Db 504 ATKVAEEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 560

Qy 53 A 53  
Db 561 A 561

RESULT 15  
US-08-346-849-16  
Sequence 16, Application US/08346849  
Patent No. 5670483  
GENERAL INFORMATION:  
APPLICANT: Zhang, Shuguang  
APPLICANT: Lockshin, Curtis  
APPLICANT: Rich, Alexander  
APPLICANT: Holmes, Todd  
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY  
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/346,849  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/973,326  
FILING DATE: 28 DECEMBER 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: MIT-6008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-346-849-16

Query Match 34.4%; Score 91.5; DB 1; Length 60;  
Best Local Similarity 52.0%; Pred. No. 0.0041;  
Matches 26; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAEKAAK-KAEAKAYKAAEAKKAEAKYKAEAAKAAK 50  
Db 1 KAAAKRKAALAKKAAKAAKRAAAKAAKAKKAKKPKKAAKAAKAAKPKK 50

Search completed: March 10, 2003, 12:30:02  
Job time : 9.92405 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	132.5	42.3	394	2	F90725	membrane spanning
2	132.5	42.3	394	2	G85576	membrane spanning
3	131	41.9	421	2	JV0057	tolA protein - Bac
4	123	39.3	248	1	HSUR1P	histone H1, gonada
5	122.5	39.1	376	2	AG0592	tolA protein [imp
6	121.5	38.8	1390	2	S51364	sperm tail-specifi
7	119.5	38.2	347	2	E83525	TolA protein PA097
8	118	37.7	231	2	S95989	histone H1 - Chlam
9	117.5	37.5	372	2	G64064	outer membrane int
10	117	37.4	388	2	AC0138	TolA colicin impor
11	115.5	36.9	206	2	S09388	histone H1 - sea u
12	115	36.7	344	2	S34153	met101-1 protein -
13	113	36.1	284	2	T06241	histone H1 (clone
14	112.5	35.9	217	2	A26721	histone H1-gamma,
15	112	35.8	243	2	AE1689	hypothetical prote
16	111.5	35.6	239	2	AE1317	hypothetical prote
17	111	35.5	219	2	A23550	histone H1 - sea u
18	110	35.1	220	2	A28456	histone H1.10 - ch
19	109	34.8	209	1	HSX11A	histone H1A - Afri
20	107	34.2	384	2	B43592	outer membrane pro
21	107	34.2	1701	2	T09127	probable erythrocy
22	106	33.9	214	2	G70673	probable hupB - My
23	105	33.5	211	2	A28100	histone H1-beta, e
24	104	33.2	311	2	T17698	hypothetical prote
25	103	32.9	182	2	S61926	histone H1 homolog
26	102.5	32.7	433	2	S25194	zuo1in - yeast (Sa
27	102	32.6	924	2	T06636	hypothetical prote
28	101.5	32.4	236	2	S23232	histone H1 - wheat
29	101.5	32.4	1128	2	T30296	R27-2 protein - Tr

A;Gene: tolA



Curr. Genet. 28, 333-345, 1995

A>Title: The organization structure and regulatory elements of Chlamydomonas histone gene  
A:Reference number: S59581; MUID:96120862; PMID:8590479

A:Accession: S59589

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-231 <FAB>

A:Cross-references: EMBL:U16726

A>Note: the authors did not translate the codon for residue 1

R:Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
submitted to the EMBL Data Library, October 1994

A>Description: The organization, structure and controlling elements of Chlamydomonas histone gene

A:Reference number: S62122

A:Accession: JG5212

A:Molecule type: DNA

A:Residues: 1-173, 'P', 174-231 <PAW>

A:Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98452.1; PID:g571480

C:Genetics:

A:Introns: 62/3; 101/3

C:Superfamily: histone H1

C:Keywords: chromosomal protein; DNA binding; nucleosome

Query Match 37.7%; Score 118; DB 2; Length 231;  
Best Local Similarity 55.6%; Pred. No. 0.009; Indels 2; Gaps 2;  
Matches 35; Conservative 5; Mismatches 21;

QY 3 KYAKKEKAYAKAKAAKAAKAAKAEAK-KYAKAAKAEKKEEYAA-AEAKYKAERAAKAAK 60  
||| ||| ||| ||| ||| ||| ||| : : : ||| |||  
Db 159 KAEXPKPAKPAKKTITTKDAAPKPAEKPPKAAKPAAKPKAAKPKAAKPAK 218  
||| ||| ||| ||| ||| ||| ||| : : : ||| |||

QY 61 EAA 63  
::

Db 219 KSA 221

RESULT 9

G64064

outer membrane integrity protein tolA - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999

C:Accession: G64064; JC5212

R:Feilichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.J.; Dwayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.G.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.

A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: G64064

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-372 <TIGR>

A:Cross-references: GB:U32722; GB:L42023; NID:g51573348; PIDN:AAC22041.1; PID:g51573353; T.R.; Sen, K.; Sikkema, D.J.; Murphy, T.F.

Gene 178, 75-81, 1996

A>Title: Isolation and characterization of the Haemophilus influenzae tolQ, tolR, tolA genes

A:Reference number: JC5212; MUID:97080550; PMID:8921895

A:Accession: JC5212

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 'Y', 2-47, 'A', 49-141, 'R', 143-164, 'P', 166-189, 'R', 191-202, 'A', 204-226, 'A', 228-247

A:Cross-references: GB:U32470

A:Experimental source: strain 1479

A>Note: the authors translated the codon CGT for residue 190 as Ala

C:Genetics:

A:Gene: tolA

A:Start codon: GTG

Query Match 37.5%; Score 117.5; DB 2; Length 372;  
Best Local Similarity 46.6%; Pred. No. 0.044; Indels 7; Gaps 1;  
Matches 34; Conservative 10; Mismatches 22;

QY 1 AKKYAKKEKAYAKAKAAKAAKAAKAEKKYAKA-----AKAEKKYYAAAEKKYAAE 53







THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 15.038 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-4

Perfect score: 313

Sequence: 1 AKYAKKEKAYAKAKBAK.....EAKYKAEAKAAKAAEAYEA 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	131	41.9	421	1	TOLA ECOLI
2	123	39.3	248	1	H1_PAKAN
3	121.5	38.8	1391	1	MST2 DROHY
4	119.5	38.2	347	1	TOLA_PSEAE
5	117.5	37.5	372	1	TOLA_HAEIN
6	115	36.7	344	1	MST1 DROHY
7	112.5	35.9	217	1	H1G_STRPU
8	111.5	35.6	208	1	DBH_MTCSM
9	111	35.5	210	1	H1_LYTPPI
10	110.5	35.3	205	1	DBH_MYCBO
11	110	35.1	219	1	H110_CHICK
12	109	34.8	209	1	H1A_XENLA
13	107	34.2	384	1	TPW8_TREPH
14	106.5	34.0	238	1	H1_WHEAT
15	106	33.9	214	1	DBH_MYCTU
16	106	33.9	349	1	R36_AEDAL
17	105	33.7	211	1	H1B_STRPU
18	102.5	32.7	433	1	ZUO1_YEAST
19	102	32.6	139	1	ASR_KLEPN
20	102	32.6	299	1	R122_DROME
21	101	32.3	206	1	H1_ONCMY
22	101	32.3	232	1	H1B_CHITE
23	101	32.3	829	1	IF2_HAEIN
24	100.5	32.1	346	1	R36_AEDAE
25	99	31.6	224	1	H1L1_CHICK
26	98.5	31.5	217	1	H1_ANAPL
27	98.5	31.5	240	1	H12_VOLCA
28	97	31.0	223	1	H103_CHICK
29	96.5	30.8	194	1	H1_SALT
30	96.5	30.8	216	1	H1C1_XENLA
31	96.5	30.8	219	1	H1B_XENLA
32	96.5	30.8	1403	1	YDF3_SCHPO
33	96	30.7	217	1	H1_CHICK

34	96	30.7	221	1	H1C_CHITE
35	95.5	30.5	220	1	H1C2_XENLA
36	95.5	30.5	232	1	H1A_CHITE
37	95.5	30.5	237	1	H1E_CHITE
38	95	30.4	1020	1	NFH_HUMAN
39	94.5	30.2	218	1	H101_CHICK
40	94.5	30.2	233	1	H11_GLYBA
41	94	30.0	273	1	H12_ARATH
42	92.5	29.6	225	1	H15_HUMAN
43	92.5	29.6	233	1	H11_GLYSA
44	92	29.4	218	1	H1E_CHICK
45	92	29.4	287	1	H1_LYCES

#### ALIGNMENTS

RESULT 1  
TOLA ECOLI  
ID TOLA ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OK NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levensgood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
their products, components of a multistep translocation system in  
Escherichia coli.";  
RT J. Bacteriol. 171:6600-6609(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Ferna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "Tola: a membrane protein involved in colicin uptake contains an  
extended helical region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
RN [5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

















GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 34.9494 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-4  
Perfect score: 313  
Sequence: 1 AKYAKKEKAYAKAKAEK.....EAKYAKAEAKAAKAEAYEA 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	42.3	394	16 Q8X965	Q8x965 escherichia
2	126.5	40.4	389	16 Q9CM70	Q9cm70 pasteurella
3	124.5	39.8	407	16 Q8ZQT6	Q8zqt6 salmonella
4	122.5	39.1	372	2 Q9WXX1	Q9wx1 pseudomonas
5	122.5	39.1	376	16 Q8ZBC1	Q8zbc1 salmonella
6	120	38.3	395	2 Q937K4	Q937k4 erwinia chr
7	117	37.4	388	16 Q8ZGZ2	Q8zgz2 versinia pe
8	113.5	36.3	275	5 O01395	O01395 drosophila
9	113	36.1	284	10 Q65794	Q65794 triticum ae
10	112	35.8	243	16 Q92A67	Q92a67 listeria mo
11	111.5	35.6	239	16 Q8Y5W4	Q8y5w4 listeria mo
12	110.5	35.3	232	10 Q39576	Q39576 chlamydomon
13	110	35.1	275	10 Q9XHL9	Q9xhl9 triticum ae
14	110	35.1	1341	16 Q98KG7	Q98kg7 rhizobium l
15	108.5	34.7	198	4 Q9H8H4	Q9h8h4 homo sapien
16	108.5	34.7	467	4 Q9H9F1	Q9h9f1 homo sapien

17	107.5	34.3	227	10 Q9SWU1	Q9swu1 triticum ae
18	107.5	34.3	238	10 Q9XHL8	Q9xhl8 triticum ae
19	107	34.2	200	16 Q8XVW7	Q8xvw7 raietonia s
20	107	34.2	1701	5 O61164	O61164 plasmodium
21	106.5	34.0	236	10 Q9SWU3	Q9swu3 triticum ae
22	106.5	34.0	237	10 Q9SWU2	Q9swu2 triticum ae
23	106.5	34.0	345	16 Q8Y1F6	Q8y1f6 raietonia s
24	106.5	34.0	755	10 Q9FP71	Q9fp71 oryza sativ
25	105	33.5	213	3 Q9UV33	Q9uv33 ascobolus i
26	104	33.2	298	2 Q52088	Q52088 pseudomonas
27	104	33.2	311	12 Q84528	Q84528 paramedum
28	104	33.2	1866	5 Q8T5C8	Q8t5c8 plasmodium
29	103	32.9	182	2 Q45370	Q45370 bordetella
30	102.5	32.7	436	11 Q99KC2	Q99kc2 mus musculu
31	102	32.6	224	13 Q90ZD7	Q90zd7 bufo bufo g
32	102	32.6	312	5 Q9UAN1	Q9uan1 drosophila
33	102	32.6	485	10 Q8RXD0	Q8rxdo arabidopsis
34	102	32.6	924	10 Q9SU08	Q9su08 arabidopsis
35	101.5	32.4	629	2 Q9KJ98	Q9kj98 escherichia
36	101.5	32.4	969	5 Q9ND19	Q9nd19 plasmodium
37	101.5	32.4	1128	5 Q26947	Q26947 trypanosoma
38	100.5	32.1	166	5 Q99281	Q99281 parechinus
39	99.5	31.8	212	3 Q93946	Q93946 candida alb
40	99.5	31.8	229	13 Q91369	Q91369 xenopus lae
41	99	31.6	581	16 Q9RU45	Q9ru45 deinococcus
42	98.5	31.5	233	5 Q18319	Q18319 chironomus
43	98.5	31.5	243	5 Q23784	Q23784 chironomus
44	98.5	31.5	607	5 Q9V6S7	Q9v6s7 drosophila
45	98	31.3	180	5 Q25636	Q25636 parechinus

#### ALIGNMENTS

##### RESULT 1

Q8X965 PRELIMINARY; PRT; 394 AA.

ID Q8X965

AC Q8X965;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Membrane spanning protein, required for outer membrane integrity

DE (Membrane spanning protein TolA).

GN TOLA OR Z0907 OR ECS0774.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OC Escherichia

OX NCBI\_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=11074935; PubMed=11206551;

RA Perna N.T., Plunkett G., Evans P.S., Gregor J., Kirkpatrick H.A., Rose D.J., Mayhew G.F., Boutin A., Shao Y., Miller L., Posfai G., Hackett J., Klink S., Lim A., Dimalanta E.T., Potamouis K., Grotbeck E.J., Davis N.W., Lin J., Yen G., Schwartz D.C., Apodaca J., Anantharaman T.S., Blattner F.R.; Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. Nature 409:529-533 (2001)."

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-T., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shingawa H.;

RT "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.," DNA Res. 8:11-22 (2001).

RL EMBL; AE005252; AAG55075.1; -

DR EMBL; AP002553; BAB34197.1; -

```
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5B58D8E8230BDE28 CRC64;

Query Match 42.3%; Score 132.5; DB 16; Length 394;
Best Local Similarity 58.9%; Pred. No. 0.00053;
Matches 43; Conservative 8; Mismatches 13; Indels 9; Gaps 4;

Qy 1 AKYAKKEKAYAKAEAKKAEAKAEAKYAKAA-----KAEKKEYAAAEAKYKAE 54
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 156 AEEAKKAADAK-KGAFAEA-KAAAEAKKAEFAAALKKKAEFAAFAAARKKAA 213
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy 55 AKAAA-KEAAYEA 66
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 214 EKAADKKAEEKA 226

RESULT 2
Q9CM70 PRELIMINARY; PRT; 389 AA.
AC Q9CM70;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TOLA.
GN TOLA OR PM0968.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.B., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006136; AAK03052.1; -.
DR HSP; P19934; ITOL.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Complete proteome.
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;

Query Match 40.4%; Score 126.5; DB 16; Length 389;
Best Local Similarity 52.0%; Pred. No. 0.0017;
Matches 39; Conservative 9; Mismatches 12; Indels 15; Gaps 3;

Qy 1 AKYAKKEKAYAKAK-----KAEAKA---AKKAEAKKYAKAKAEKKEYAAAEAKYK 51
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 189 AKRKAERAKAEAEAKKAEKAEAEKVKAEKAEAEKAEAEAKAEAKAEAK 242
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy 52 ABAKAAKAEAYEA 66
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 243 AEKAKADAEAAQRKA 257

RESULT 3
Q8ZQT6 PRELIMINARY; PRT; 407 AA.
AC Q8ZQT6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tol protein, membrane spanning protein.
GN TOLA OR STM0747.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
DR EMBL; AE008730; AAL19691.1; -.
DR InterPro; IPR00104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Complete proteome.
SQ SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;

Query Match 39.8%; Score 124.5; DB 16; Length 407;
Best Local Similarity 47.8%; Pred. No. 0.0026;
Matches 43; Conservative 6; Mismatches 10; Indels 31; Gaps 4;

Qy 1 AKYAKKEKAYAKA---KKAEEAKAK-----KAKAEAKKYAKAKAEKKEY 43
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 189 AKKAEAEAKKAAAEAKKAEAEAKKAAAEAKKADAEAKKADAEAKKADAEAKK 244
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Qy 44 AAAEAKYKAEAA-----KAAAKEAA 63
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 245 AAAEAKKADAAAKAAADAKKKKAAAEKAA 274

RESULT 4
Q9WMX1 PRELIMINARY; PRT; 372 AA.
AC Q9WMX1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Tola protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX MEDLINE=96198174; PubMed=8626299;
RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
RT lipoprotein (PAL) is involved in maintenance of the integrity of the
RT cell envelope.";
RL J. Bacteriol. 178:1699-1706 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX Ramos-Gonzalez I.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX Rodriguez-Herva J.J.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX MEDLINE=96422022; PubMed=8824639;
RA Rodriguez-Herva J.J., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RL J. Bacteriol. 178:5836-5840 (1996).
RN [5]
RP EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;
```

```

Query Match          39.1%; Score 122.5; DB 2; Length 372;
Best Local Similarity 51.5%; Pred. NO. 0.0035;
Matches 35; Conservative 10; Mismatches 18; Indels 5; Gaps 2;

Qy      1 AKKYAKKEKAYAKAKAAKAAGAAGKAKAAGKAYAKAAKAEKKYAAAEKYYAAAEAKYKA--EAAKAA 58
Db      148 AKKAAGQAADIAXKKADEAKKAEAKKEEEK---KAAAEAKKAAAEADAKKAAAEAKKKA 204
Qy      59 AKEAYEEA 66
Db      205 AEDAKKA 212

RESULT 5
Q8Z8C1 PRELIMINARY; PRT; 376 AA.
AC Q8Z8C1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tola protein.
GN STY0793.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=15134947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks R.K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627288; CAD05209.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZEI.
DR PRINTS; PR01574; TUBBYPROTEIN.
KW Complete proteome.
SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;

Query Match          39.1%; Score 122.5; DB 16; Length 376;
Best Local Similarity 51.3%; Pred. NO. 0.0036;
Matches 40; Conservative 4; Mismatches 13; Indels 21; Gaps 3;

Qy      1 AKKYAKKEKAYAKA---KKAENKAAK-----KAKBAKYYAKAAKAEKEY 43
Db      173 AKKAAEAFAAKAAAEAKKKAERAKAAKAAADAKKADAFAAKAAAEAKCKDAATAAK---- 228
Qy      44 AAAEAKYKAAAKAAAKE 61
Db      229 AAADAKKAAAEKAAANAE 246

RESULT 6
Q937K4 PRELIMINARY; PRT; 395 AA.
ID Q937K4
AC Q937K4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tola protein.
GN TOLA.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
```

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE Axoneme-associated protein MST101(3).  
 GN MST101(3) OR DHMST101.  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Neesen J., Heinlein U.A.O., Buenemann H.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).  
 CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
 DR EMBL; U85627; AAB51369.1; -.  
 DR FlyBase; FBgn020732; Dhyd\mst101(3).  
 KW Sperm; Repeat; Multigene family.  
 FT DOMAIN 64 255  
 FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-  
 FT [AE]-X.  
 SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2A2DF732C CRC64;  
 Query Match 36.3%; Score 113.5; DB 5; Length 275;  
 Best Local Similarity 52.7%; Pred. No. 0.015;  
 Matches 39; Conservative 3; Mismatches 15; Indels 17; Gaps 4;  
 QY 5 AKKEKAYAKKAKAKAKAK--AEAKYAKAKAE-----KKEYAAAEAK 49  
 DB 71 AKKEKAAKKKC-AEAAKKEKAAEKKCAEAAKKEQAAQKKCAELAKKEKAAEK 129  
 QY 50 YKAAAKAAKAAEAA 63  
 DB 130 KCAEAAK-KEKEAA 142  
 RESULT 9  
 ID O65794 PRELIMINARY; PRT; 284 AA.  
 AC O65794;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Histone H1.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98249625; PubMed=9588026;  
 RA Taoka K., Ohtsubo N., Fujimoto Y., Mikami K., Meshi T., Iwabuchi M.;  
 RT "The modular structure and function of the wheat H1 promoter with S  
 RT phase-specific activity";  
 RL Plant Cell Physiol. 39:294-306(1998).  
 DR EMBL; D87064; BAA25203.1; -.  
 DR HSP; P02259; 1HST.  
 DR InterPro; IPR000637; AT hook.  
 DR InterPro; IPR001386; Histone H1/H5.  
 DR InterPro; IPR003216; LinkerHist\_N.  
 DR Pfam; PF02178; AT hook; 1.  
 DR Pfam; PF00538; linker histone; 1.  
 DR ProDom; PD000373; Linkerhist\_N; 1.  
 DR SMART; SM00384; AT hook; 1.  
 DR SMART; SM00526; H15; 1.  
 SQ SEQUENCE 284 AA; 29159 MW; 5165A2B922BCBA4E CRC64;  
 Query Match 36.1%; Score 113; DB 10; Length 284;  
 Best Local Similarity 52.5%; Pred. No. 0.017;

Matches 32; Conservative 6; Mismatches 21; Indels 2; Gaps 1;  
 QY 2 KKYAKKEKAYAKKAAKAAK--AKAAKAYAKAAKAEKKEYAAAEAKYKAAEAKAA 59  
 DB 152 KPPAAKKKAPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 211  
 QY 60 K 60  
 DB 212 K 212  
 RESULT 10  
 ID Q92A67 PRELIMINARY; PRT; 243 AA.  
 AC Q92A67;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein lin2055.  
 GN LIN2055.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Chabrit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo P., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunsch F., Kurapkai G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;  
 RT "Comparative genomics of Listeria species";  
 RL Science 294:849-852(2001).  
 DR EMBL; AL596170; CAC97285.1; -.  
 DR ListiList; LIN02055; -.  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 243 AA; 25963 MW; 6B2493D143B159D1 CRC64;  
 Query Match 35.8%; Score 112; DB 16; Length 243;  
 Best Local Similarity 47.1%; Pred. No. 0.018;  
 Matches 32; Conservative 11; Mismatches 23; Indels 2; Gaps 1;  
 QY 1 AKKYAKKEKAYAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 58  
 DB 110 ABEKAAAEKAAEKEKAAEKEKAAEKEKAAEKEKAAEKEKAAEKEKAAEKEKAA 169  
 QY 59 AKEAAVEA 66  
 DB 170 KEKAAEA 177  
 RESULT 11  
 ID Q8Y5W4 PRELIMINARY; PRT; 239 AA.  
 AC Q8Y5W4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein lmo1941.  
 GN LMO1941.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.



```
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002997; BAB48847.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1341 AA; 138670 MW; 79C5D8149C969BC8 CRC64;

Query Match 35.1%; Score 110; DB 16; Length 1341;
Best Local Similarity 53.5%; Pred. No. 0.14;
Matches 38; Conservative 9; Mismatches 16; Indels 8; Gaps 4;

Qy 1 AKKYAKKE---KAYAKAK-KAEAKAAKAKAEAKKYAKA---AKAEKKEYAAAEAKKYKAE 53
Db 1195 AKAKAEAEQAQAQAEAKQAQAEAKAKADAEAKAEAEAKAKADAEAKADAEAKAKAD 1254
Qy 54 A-AKAAAKEAA 63
Db 1255 ADAKAAAEQQA 1265

RESULT 15
Q9H8H4 PRELIMINARY; PRT; 198 AA.
AC Q9H8H4
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ13629 fis, clone PLACE1011056, weakly similar to histone H1,
DS gonadal.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Iehii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RA "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK023691; BAB14642.1; -.
DR InterPro; IPR000637; AT hook.
DR SMART; SM00384; AT hook; 1.
SQ SEQUENCE 198 AA; 21353 MW; 46F43602BC12487B CRC64;

Query Match 34.7%; Score 108.5; DB 4; Length 198;
Best Local Similarity 50.7%; Pred. No. 0.029;
Matches 36; Conservative 9; Mismatches 17; Indels 9; Gaps 4;

Qy 1 AKKYAKKEKAYAKAKAEAKAK---KAKA---EAKKYAKAA--KAEKKEYAR-AEAKYK 51
Db 17 ARTLAKARQAQAKVARTQAQAKAKAKAKAALVKAKAKAKAAQVYKAKAKVMAARAKAKAK 76
Qy 52 AEAATAAAKEA 62
Db 77 AKAVRAKAKVA 87

Search completed: March 10, 2003, 12:25:56
Job time : 35.9494 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 34.3924 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989a-4

Perfect score: 313

Sequence: 1 AKYAKKAYAKAKAEAK.....EAKYKAEAKAAKAEAYEA 66

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	313	100.0	66	21	AA082574
2	282.5	90.3	77	21	AA082575
3	268	85.6	86	21	AA082576
4	228.5	73.0	109	21	AA082577
5	198	63.3	56	21	AA082573
6	129.5	41.4	154	11	AA06445
7	128	40.9	100	21	AA08499
8	128	40.9	100	21	AA05904
9	128	40.9	100	22	AA04289
10	128	40.9	100	22	AA04582

11	122	39.0	106	11	AA06446	Recombinant copoly
12	120.5	38.5	45	21	AA082572	Copolymer molecule
13	111.5	35.6	239	23	AB049123	Listeria monocytog
14	110.5	35.3	205	21	AB020575	Mycobacterium bovi
15	108.5	34.7	198	22	AA095499	Human protein sequ
16	108.5	34.7	279	22	AA003592	Human DNA modifia
17	108.5	34.7	467	22	AA094309	Human protein sequ
18	106	33.9	214	20	AA034055	M. tuberculosis hi
19	106	33.9	214	21	AA057353	M. tuberculosis hi
20	105.5	33.7	35	21	AA082571	Copolymer molecule
21	105	33.5	1507	21	AA024128	Plasmodium Yoelii
22	104.5	33.4	223	20	AA014928	Amino acid sequenc
23	104	33.2	140	13	AA025206	Synthetic helical
24	103	32.9	334	22	AB028693	Novel human diagno
25	102.5	32.7	433	18	AA030256	Zuotin. Saccharom
26	102.5	32.7	433	20	AA032954	Human zuotin prote
27	102.5	32.7	433	22	AA070739	S cerevisiae apopt
28	102	32.6	46	18	AA024449	Nucleic acid (NA)
29	102	32.6	299	22	AB063276	Drosophila melanog
30	101.5	32.4	643	16	AA084568	Trypanosoma cruzi
31	101	32.3	829	22	AA035623	Haemophilus influe
32	100.5	32.1	564	16	AA084565	Trypanosoma cruzi
33	99	31.6	859	23	AA080266	Bacillus sp alkali
34	98.5	31.5	607	22	AB062173	Drosophila melanog
35	98	31.3	219	21	AA076981	Quail H1 histone p
36	95.5	30.5	472	16	AA084569	Trypanosoma cruzi
37	95	30.4	160	18	AA014549	Streptococcus pneu
38	95	30.4	427	22	AA070868	C albicans apoptos
39	95	30.4	617	22	AB035445	Peptide #2951 enco
40	95	30.4	617	22	AB020884	Protein #2883 enco
41	95	30.4	617	22	AA056269	Human brain expres
42	95	30.4	617	22	AA016458	Peptide #2892 enco
43	95	30.4	617	22	AA04187	Peptide #2869 enco
44	95	30.4	1026	22	AA078825	Human protein SEQ
45	94	30.0	46	13	AA028871	High affinity macr

#### ALIGNMENTS

##### RESULT 1

AA082574

ID AA082574 standard; peptide; 66 AA.

XX AA082574;

XX AC

DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antiinflammatory; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX WO200018794-A1.

XX 06-APR-2000.

XX 24-SEP-1999; 99WO-US22402.

XX 25-SEP-1998; 98US-0101693.

XX (YEDA ) YEDA RES & DEV CO LTD.

XX (TEVA-) TEVA PHARM USA INC.

```
XX Gad A, Lis D;
PI WPI; 2000-317499/27.
XX Copolymer 1 related polypeptides used as molecular weight markers for
DR glatiramer acetate and for treatment and prevention of immune diseases
XX Claim 10; Page 14; 72pp; English.
XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, psoriasis, pemphigus vulgaris, or systemic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX SQ Sequence 66 AA;
Query Match 100.0%; Score 313; DB 21; Length 66;
Best Local Similarity 100.0%; Pred. No. 3e-24;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKKYAKKEKAYAKAEAKKAAKAEAKKAAKAEKKEKAYAAAEKKEKAYAAAEK 60
Db 1 AKKYAKKEKAYAKAEAKKAAKAEAKKAAKAEKKEKAYAAAEKKEKAYAAAEK 60
Qy 61 EAYEA 66
Db 61 EAYEA 66
RESULT 2
AAY82575
ID AAY82575 standard; peptide; 77 AA.
XX AC AAY82575;
XX DT 28-JUL-2000 (first entry)
XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.
XX Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX Unidentified.
OS
XX WO200018794-A1.
PN
```

```
XX 06-APR-2000.
XX 24-SEP-1999; 99WO-US22402.
XX 25-SEP-1999; 98US-0101693.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX (TEVA-) TEVA PHARM USA INC.
XX Gad A, Lis D;
XX WPI; 2000-317499/27.
XX Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX Claim 10; Page 14; 72pp; English.
XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, psoriasis, pemphigus vulgaris, or systemic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX SQ Sequence 77 AA;
Query Match 90.3%; Score 282.5; DB 21; Length 77;
Best Local Similarity 84.4%; Pred. No. 3.6e-21;
Matches 65; Conservative 0; Mismatches 1; Indels 11; Gaps 2;
Qy 1 AKKYAKKEKAYAKAEAKKAAKAEAKKAAKAEKKEKAYAAAEK 49
Db 1 AKKYAKKEKAYAKAEAKKAAKAEAKKAAKAEKKEKAYAAAEK 60
Qy 50 YKAEAAKAAAEKAEAYEA 66
Db 61 YKAEAAKAAAEKAEAYEA 77
RESULT 3
AAY82576
ID AAY82576 standard; peptide; 86 AA.
XX AC AAY82576;
XX DT 28-JUL-2000 (first entry)
XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
XX Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.
XX Unidentified.
OS
XX WO200018794-A1.
PN
```



Db 61 YKAEAKKYAKAAKAEKKEYYAAAEAKKAAKAYKAEAAKAAKAEAAEYEA 109

## RESULT 5

AY82573  
ID AAY82573 standard; peptide; 56 AA.

XX AAY82573;

XX AC

XX DT 28-JUL-2000 (first entry)

XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.

XX OS Unidentified.

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX FA (TEVA-) TEVA PHARM USA INC.

XX PI Gad A, Lis D;

XX DR WPI; 2000-317499/27.

XX PT Copolymer 1 related polypeptides used as molecular weight markers for  
PT glatiramer acetate and for treatment and prevention of immune diseases

XX PS Claim 10; Page 14; 72pp; English.

XX CC AY82571 to AY82577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular  
CC weight of a copolymer (CP), which has an identified molecular weight  
CC and an amino acid composition corresponding to the copolymer. The  
CC polypeptides of the invention are used as molecular weight markers for  
CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
CC used for treating and preventing immune diseases in a mammal. Autoimmune  
CC diseases which may be treated include either cell-mediated or  
CC antibody-mediated diseases. Such diseases include arthritic conditions,  
CC demyelinating diseases and inflammatory conditions, e.g. multiple  
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
CC lupus erythematosus. Mediated-mediated diseases which can be treated  
CC include host-versus-graft disease, graft-versus-host disease, and  
CC delayed-type hypersensitivity. The polypeptides of the invention have  
CC defined molecular weights and physical properties which are analogous to  
CC glatiramer acetate molecules, which makes them ideal for use as  
CC molecular weight markers.

XX Sequence 56 AA;

Query Match 63.3%; Score 198; DB 21; Length 56;

Best Local Similarity 80.3%; Pred. No. 5.9e-13;  
Matches 53; Conservative 2; Mismatches 1; Indels 10; Gaps 5;

OY 1 AKKYAKKCKYAKAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 60  
|||||  
DB 1 AKKYAKKCKYAKAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 50  
|||||

OY 61 EAAEYA 66  
|||||

DB 51 EAAEYA 56  
|||||

## RESULT 6

AAR06445

ID AAR06445 standard; protein; 154 AA.

XX AC AAR06445;

XX DT 03-JAN-1991 (first entry)

XX DE Recombinant copolymer 1-77, myelin basic protein analogue.

XX KW Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;

XX KW immunological activity; autoimmune encephalomyelitis;

XX KW multiple sclerosis;

XX OS Synthetic.

XX PN EP383620-A.

XX PD 22-AUG-1990.

XX PF 16-FEB-1990; 90EP-0301700.

XX PR 07-FEB-1990; 90US-0473845.

XX PR 17-FEB-1989; 89US-0312541.

XX PA (REPL-) REPLIGEN CORP.

XX PI Cook KS;

XX DR WPI; 1990-255848/34.

XX DR N-PSDB; AAQ05664.

XX PT Producing genes encoding random polymers of aminoacid(s) - for  
PT producing recombinant polypeptide(s) with biological and/or  
PT immunological activity

XX Disclosure; Fig 11; 25pp; English.

XX CC To improve the expression of rCOP-1 polypeptides in E. coli, genes  
CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN  
CC (deposited: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
CC express Protein A. The resulting plasmids encode fusion proteins  
CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
CC A methionine residue occurs between the Protein A and rCOP-1  
CC sequences, originating from the 5' linker sequence, in order that  
CC the COP-1 polypeptide may be cleaved from the fusion protein.  
CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
CC segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue  
CC is left behind following CNBr cleavage of the fusion protein.  
CC The product prevents or arrests experimental autoimmune  
CC encephalomyelitis. They are used to prevent, arrest or control a  
CC demyelinating disorder, e.g. multiple sclerosis. They may also  
CC be used as additives to hair care products to confer beneficial  
CC effects on damaged hair or as supplements for diets deficient in  
CC certain amino acids.

XX See also AAQ05665.

XX Sequence 154 AA;

Query Match 41.4%; Score 129.5; DB 11; Length 154;  
Best Local Similarity 48.8%; Pred. No. 1e-05;







Best Local Similarity 71.1%; Pred. No. 2.2e-05;  
Matches 32; Conservative 0; Mismatches 6; Indels 7; Gaps 2;

QY 29 AKKYAKAAKAE--KKEYAAAAEK-----YKAEAAKAAKAAKEAAYEA 66  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 1 AKKYAKKAERAKKAYKAAEAKKAAYKAAEKAAKAAKEAAYEA 45

RESULT 13  
ABBA49123  
ID ABB49123 standard; Protein; 239 AA.  
AC ABB49123;  
XX XX  
DT 05-FEB-2002 (first entry)  
XX XX  
DE Listeria monocytogenes protein #1827.  
XX XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX XX  
OS Listeria monocytogenes.  
XX XX  
PN WO200177335-A2.  
XX XX  
PD 18-OCT-2001.  
XX XX  
PP 11-APR-2001; 2001WO-FR01118.  
XX XX  
PR 11-APR-2000; 2000FR-0004629.  
XX XX  
PA (INSP.) INST PASTEUR.  
XX XX  
PI Buchrieser C, Frangeul L, Couve E, Rueniock C, Fsihi H, Deboux P;  
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst P, Cossart P;  
PI Daniels J, Goebel W, Krieff J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Madueno E, De Pablo B, Wehlend J, Kaerst U, Entian K, Hauf J;  
XX Rose M, Voss H;  
XX MPI; 2002-010914/01.  
DR XX  
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and  
PT related polypeptides -  
XX XX  
PS Claim 6; SEQ ID No 1828; 192pp; French.  
XX XX  
XX The present invention relates to the genome sequence of *Listeria*  
CC monocytes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in *L.*  
CC monocytes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC expressed from the genome sequence of the present invention. Proteins  
CC antibodies, identification of *L. monocytogenes* and related specific  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by *L.*  
CC monocytes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX XX  
SQ Sequence 239 AA;  
  
Query Match 35.6%; Score 111.5; DB 23; Length 239;  
Best Local Similarity 46.4%; Pred. No. 0.001;





**THIS PAGE BLANK (USPTO)**







[illegible]



THIS PAGE BLANK (USPTO)















```
Best Local Similarity 46.0%; Pred. No. 0.0099;
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAAK 60
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390
Qy 61 EAA 63
Db 391 EAA 393

RESULT 14
US-09-115-746-2
Sequence 2, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-2

Query Match 32.4%; Score 101.5; DB 4; Length 564;
Best Local Similarity 46.0%; Pred. No. 0.0099;
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAAK 60
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390
Qy 61 EAA 63
Db 391 EAA 393

RESULT 15
US-08-216-894-8
Sequence 8, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-8

Query Match 32.4%; Score 101.5; DB 2; Length 643;
Best Local Similarity 46.0%; Pred. No. 0.011;
Matches 29; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 2 KKYAKKEKAYAKAKAEAKKAKAEAKKYAKAKA-EKKEYAAAEAKKYAKAAK 60
Db 331 QRAAEAAKAVETEKQRAAEATKVAAEKRAAEAKAVETEKQRAAEATKVAAEKQKAA 390
Qy 61 EAA 63
Db 391 EAA 393

Search completed: March 10, 2003, 12:30:03
Job time : 12.6962 secs
```

THIS PAGE BLANK (USPTO)



F90725



A: Molecule type: DNA  
A: Residues: 'V', '2', '47', 'A', '49-141', 'R', '143-164', 'P', '166-189', 'R', '191-202', 'A', '204-226', 'A', '228-247', 'A', '250-259', 'A', '261-269', 'A', '271-279', 'A', '281-289', 'A', '291-299', 'A', '301-309', 'A', '311-319', 'A', '321-329', 'A', '331-339', 'A', '341-349', 'A', '351-359', 'A', '361-369', 'A', '371-379', 'A', '381-389', 'A', '391-399', 'A', '401-409', 'A', '411-419', 'A', '421-429', 'A', '431-439', 'A', '441-449', 'A', '451-459', 'A', '461-469', 'A', '471-479', 'A', '481-489', 'A', '491-499', 'A', '501-509', 'A', '511-519', 'A', '521-529', 'A', '531-539', 'A', '541-549', 'A', '551-559', 'A', '561-569', 'A', '571-579', 'A', '581-589', 'A', '591-599', 'A', '601-609', 'A', '611-619', 'A', '621-629', 'A', '631-639', 'A', '641-649', 'A', '651-659', 'A', '661-669', 'A', '671-679', 'A', '681-689', 'A', '691-699', 'A', '701-709', 'A', '711-719', 'A', '721-729', 'A', '731-739', 'A', '741-749', 'A', '751-759', 'A', '761-769', 'A', '771-779', 'A', '781-789', 'A', '791-799', 'A', '801-809', 'A', '811-819', 'A', '821-829', 'A', '831-839', 'A', '841-849', 'A', '851-859', 'A', '861-869', 'A', '871-879', 'A', '881-889', 'A', '891-899', 'A', '901-909', 'A', '911-919', 'A', '921-929', 'A', '931-939', 'A', '941-949', 'A', '951-959', 'A', '961-969', 'A', '971-979', 'A', '981-989', 'A', '991-999', 'A', '1001-1009', 'A', '1011-1019', 'A', '1021-1029', 'A', '1031-1039', 'A', '1041-1049', 'A', '1051-1059', 'A', '1061-1069', 'A', '1071-1079', 'A', '1081-1089', 'A', '1091-1099', 'A', '1101-1109', 'A', '1111-1119', 'A', '1121-1129', 'A', '1131-1139', 'A', '1141-1149', 'A', '1151-1159', 'A', '1161-1169', 'A', '1171-1179', 'A', '1181-1189', 'A', '1191-1199', 'A', '1201-1209', 'A', '1211-1219', 'A', '1221-1229', 'A', '1231-1239', 'A', '1241-1249', 'A', '1251-1259', 'A', '1261-1269', 'A', '1271-1279', 'A', '1281-1289', 'A', '1291-1299', 'A', '1301-1309', 'A', '1311-1319', 'A', '1321-1329', 'A', '1331-1339', 'A', '1341-1349', 'A', '1351-1359', 'A', '1361-1369', 'A', '1371-1379', 'A', '1381-1389', 'A', '1391-1399', 'A', '1401-1409', 'A', '1411-1419', 'A', '1421-1429', 'A', '1431-1439', 'A', '1441-1449', 'A', '1451-1459', 'A', '1461-1469', 'A', '1471-1479', 'A', '1481-1489', 'A', '1491-1499', 'A', '1501-1509', 'A', '1511-1519', 'A', '1521-1529', 'A', '1531-1539', 'A', '1541-1549', 'A', '1551-1559', 'A', '1561-1569', 'A', '1571-1579', 'A', '1581-1589', 'A', '1591-1599', 'A', '1601-1609', 'A', '1611-1619', 'A', '1621-1629', 'A', '1631-1639', 'A', '1641-1649', 'A', '1651-1659', 'A', '1661-1669', 'A', '1671-1679', 'A', '1681-1689', 'A', '1691-1699', 'A', '1701-1709', 'A', '1711-1719', 'A', '1721-1729', 'A', '1731-1739', 'A', '1741-1749', 'A', '1751-1759', 'A', '1761-1769', 'A', '1771-1779', 'A', '1781-1789', 'A', '1791-1799', 'A', '1801-1809', 'A', '1811-1819', 'A', '1821-1829', 'A', '1831-1839', 'A', '1841-1849', 'A', '1851-1859', 'A', '1861-1869', 'A', '1871-1879', 'A', '1881-1889', 'A', '1891-1899', 'A', '1901-1909', 'A', '1911-1919', 'A', '1921-1929', 'A', '1931-1939', 'A', '1941-1949', 'A', '1951-1959', 'A', '1961-1969', 'A', '1971-1979', 'A', '1981-1989', 'A', '1991-1999', 'A', '2001-2009', 'A', '2011-2019', 'A', '2021-2029', 'A', '2031-2039', 'A', '2041-2049', 'A', '2051-2059', 'A', '2061-2069', 'A', '2071-2079', 'A', '2081-2089', 'A', '2091-2099', 'A', '2101-2109', 'A', '2111-2119', 'A', '2121-2129', 'A', '2131-2139', 'A', '2141-2149', 'A', '2151-2159', 'A', '2161-2169', 'A', '2171-2179', 'A', '2181-2189', 'A', '2191-2199', 'A', '2201-2209', 'A', '2211-2219', 'A', '2221-2229', 'A', '2231-2239', 'A', '2241-2249', 'A', '2251-2259', 'A', '2261-2269', 'A', '2271-2279', 'A', '2281-2289', 'A', '2291-2299', 'A', '2301-2309', 'A', '2311-2319', 'A', '2321-2329', 'A', '2331-2339', 'A', '2341-2349', 'A', '2351-2359', 'A', '2361-2369', 'A', '2371-2379', 'A', '2381-2389', 'A', '2391-2399', 'A', '2401-2409', 'A', '2411-2419', 'A', '2421-2429', 'A', '2431-2439', 'A', '2441-2449', 'A', '2451-2459', 'A', '2461-2469', 'A', '2471-2479', 'A', '2481-2489', 'A', '2491-2499', 'A', '2501-2509', 'A', '2511-2519', 'A', '2521-2529', 'A', '2531-2539', 'A', '2541-2549', 'A', '2551-2559', 'A', '2561-2569', 'A', '2571-2579', 'A', '2581-2589', 'A', '2591-2599', 'A', '2601-2609', 'A', '2611-2619', 'A', '2621-2629', 'A', '2631-2639', 'A', '2641-2649', 'A', '2651-2659', 'A', '2661-2669', 'A', '2671-2679', 'A', '2681-2689', 'A', '2691-2699', 'A', '2701-2709', 'A', '2711-2719', 'A', '2721-2729', 'A', '2731-2739', 'A', '2741-2749', 'A', '2751-2759', 'A', '2761-2769', 'A', '2771-2779', 'A', '2781-2789', 'A', '2791-2799', 'A', '2801-2809', 'A', '2811-2819', 'A', '2821-2829', 'A', '2831-2839', 'A', '2841-2849', 'A', '2851-2859', 'A', '2861-2869', 'A', '2871-2879', 'A', '2881-2889', 'A', '2891-2899', 'A', '2901-2909', 'A', '2911-2919', 'A', '2921-2929', 'A', '2931-2939', 'A', '2941-2949', 'A', '2951-2959', 'A', '2961-2969', 'A', '2971-2979', 'A', '2981-2989', 'A', '2991-2999', 'A', '3001-3009', 'A', '3011-3019', 'A', '3021-3029', 'A', '3031-3039', 'A', '3041-3049', 'A', '3051-3059', 'A', '3061-3069', 'A', '3071-3079', 'A', '3081-3089', 'A', '3091-3099', 'A', '3101-3109', 'A', '3111-3119', 'A', '3121-3129', 'A', '3131-3139', 'A', '3141-3149', 'A', '3151-3159', 'A', '3161-3169', 'A', '3171-3179', 'A', '3181-3189', 'A', '3191-3199', 'A', '3201-3209', 'A', '3

QY	60	KYKAEAAKAAKAAA	74
Db	156	PAKKAACKYPAAKKAA	170

RESULT 9  
E60110  
repetitive protein antigen 27 - Trypanosoma cruzi (fragments)  
C:Species: Trypanosoma cruzi  
C:Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 02-Jun-2000  
C:Accession: E60110  
R:Hoft D.F.; Kim K.S.; Otsu K.; Moser D.R.; Vost W.J.; Blumin J.H.; Donelson J.E.

A;Title: Myxoplasma ciliar expresses diverse repetitive protein antigens.  
A;Reference number: A60110; MUID:89277508; PMID:2659529  
A;Accession: E60110  
A;Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A;Residues: 1-219 <HOP>  
C;Comment: This protein contains a series of tandem repeats, each fourteen residues in length.  
C;Superfamily: neurofilament triplet H protein  
C;Keywords: tandem repeat

Query Match            36.7%;    Score 134.5; DB 2; Length 219;  
Best Local Similarity   50.7%;    Pred.No.0.0018;  
Matches   38; Conservative   8; Mismatches   28; Indels   1; Gaps   1;

Qy         1   AKTYAKKEKAYAKAEKAACKAEAKYKAAEAKKKAKAEAKKYAKAAK-AEKKEYAAAAEA   59  
            ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db         70   ATKVAGDEKQAAEATKVYKVAEAEKQKQAEATKVYAEAEKQKQAEATKVYAEAEKQKQAAA   129  
            ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy         60   KYKAEAKAAKAAKEAA    74  
            ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Db         130   TKVAEAGKQKAAEEA     144  
            ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 10  
S59589

C;Species: Chlamydomonas reinhardtii  
 C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 23-Jul-1999  
 C;Accession: S59589; S62122

**A:**Title: The organization structure and regulatory elements of Chlamydomonas histone gene A;Reference number: S59581; MUID:96120862; PMID:8590479  
Curr. Genet. 28, 333-345, 1995

A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-231 <FAB>  
A;Cross-references: EMBL:U16726  
A;Note: the authors did not translate the codon for residue 1  
R;Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
submitted to the EMBL Data Library. October 1994

Query Match	36.3%;	Score 133;	DB 2;	Length 328;
Best Local Similarity	52.7%;	Pred. No. 0.0031;		

A/Accession: S31364  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-1390 <NEE>  
A/Cross-references: EMBL:X73481  
R/Reesen, J.; Heinlein, U.A.O.; Buenemann, H.  
submitted to the EMBL Data Library, June 1993  
A/Reference number: S34153  
A/Accession: S34154  
A/Molecule type: DNA  
A/Residues: 1-163, 'E', 164-236, 'Q', 237-354, 257-320, 'E', 321-1390 <NEW>  
A/Cross-references: EMBL:X73481; NID:g313201; PID:g313202  
C/Genetics:  
A/Gene: mat101 (2)  
A/Cross-references: FlyBase:FBgn0011816

Query Match	35.5%;	Score 130;	DB 2;	Length 1390;
Best Local Similarity	51.7%;	Pred. No. 0.015;		

Matches 45; Conservative 3; Mismatches 23; Indels 16; Gaps 5;

Qy 2 KKYAKKEK--AYAKAEAKA--AEAKAYAAEAKKAKAEAKKYAKAAK-----AE 50

Db 511 KKLAKKEKETAEKKCEKAAKKEAEKKCEKAAKKEAEKKCEKAAKKEAEKKCEKAAKKEAE 570

Qy 51 KK--EYAAAEAKYKAEAAK---AAAKE 72

Db 571 KKKCEKAAKERKAEAAEKKKCEAEAAKKE 597

## RESULT 15

G70673

probable hupB - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C;Accession: G70673

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: G70673

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-214 &lt;COL&gt;

A;Cross-references: GB:283018; GB:AL123456; NID:g3261671; PIDN:CAB05427.1; PID:g1694845

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: hupB

C;Superfamily: histone H1

## Query Match

35.4%; Score 129.5; DB 2; Length 214;

Best Local Similarity 45.7%; Pred. No. 0.004;

Matches 43; Conservative 4; Mismatches 28; Indels 19; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAEAK-----AYKAAEAKKKAKAEAKKYAKAAKAEKKEY 54

Db 111 AKKVAK--KAPAKKATKAACKAATKAPAKKAATKAPAKKAATKAPAKKAVKATKSPAKKV 168

Qy 55 AAA-----EAKYKAEAAKAAAEAAVEA 77

Db 169 TKAVKKTAVKASVRKAATKAPAKKAAAKRPATKA 202

Search completed: March 10, 2003, 12:28:28

Job time : 23.5802 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 17.5443 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-5

Perfect score: 366

Sequence: 1 AKYAKKEKAYAKKAERAAK.....EAKYAKBAERAAKAAERAYEA 77

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150.5	41.1	421	1 TOLA_ECOLI	P19934 escherichia
2	141.5	38.7	347	1 TOLA_PSEAE	P50600 pseudomonas
3	141	38.5	372	1 TOLA_HAEIN	P44678 haemophilus
4	136	37.2	210	1 H1_LYTP1	P06144 lytechinus
5	130	35.5	205	1 DBH_MYCBO	Q9xb18 mycobacteri
6	130	35.5	1391	1 MST2_DROHY	Q08696 drosophila
7	129.5	35.4	214	1 DBH_WCTU	P95109 mycobacteri
8	129	35.2	208	1 DBH_MTCSM	Q9zhc5 mycobacteri
9	127.5	34.8	344	1 MST1_DROHY	Q08695 drosophila
10	127	34.7	217	1 H1G_STRPU	P07796 strongyloce
11	125.5	34.3	211	1 H1B_STRPU	P15869 strongyloce
12	125.5	34.3	349	1 RS6_AEDAL	Q9u762 aedes albop
13	123	33.6	248	1 H1_PARAN	P02256 paracitrus
14	123	33.6	346	1 RS6_AEDAE	Q9u761 aedes aegypt
15	123	33.6	433	1 ZUO1_YEAST	P32527 saccharomyc
16	122.5	33.5	219	1 H110_CHICK	P08286 gallus gall
17	121.5	33.2	384	1 TWBP_TREPH	P23720 treponema p
18	120.5	32.9	221	1 H1C_CHITE	P40277 chironomus
19	118.5	32.4	240	1 H12_VOLCA	Q08865 volvox cart
20	118	32.2	206	1 H1_ONCMY	P06350 oncothrychnu
21	118	32.2	237	1 H1E_CHITE	P40278 chironomus
22	118	32.2	1403	1 YDF3_SCHPO	Q10475 schizosacch
23	115.5	31.6	224	1 H11L_CHICK	P08287 gallus gall
24	115	31.4	232	1 H1B_CHITE	P40276 chironomus
25	114	31.1	233	1 H1L_GLYSA	P40266 glyptotendi
26	113.5	31.0	223	1 H103_CHICK	P08285 gallus gall
27	113.5	31.0	917	1 IF2_PROVU	Q9zf22 proteus vul
28	113	30.9	238	1 H1_WHEAT	P27806 triticum ae
29	112.5	30.7	217	1 H1_CHICK	P09987 gallus gall
30	112	30.6	217	1 H1_ANAPL	P09426 anas platyr
31	111.5	30.5	139	1 ASR_KLEPN	Q93mh5 klebsiella
32	111.5	30.5	325	1 TWBP_TREPA	P19649 treponema p
33	111	30.3	207	1 H11_CAEEI	P10771 caenorhabdi

RESULT 1					
TOLA_ECOLI					
ID	TOLA_ECOLI	STANDARD;	PRT;	421 AA.	
AC	P19934;				
DT	01-FEB-1991 (Rel. 17, Created)				
DT	01-FEB-1991 (Rel. 17, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	TOLA protein.				
GN	TOLA OR CIM OR EXCC OR LKY OR B0739.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=JM105;				
EX	MEDLINE=90078104; PubMed=2687247;				
RA	Levensgood S.K., Webster R.E.;				
RT	"Nucleotide sequences of the tola and tolb genes and localization of				
RT	their products, components of a multistep translocation system in				
RT	Escherichia coli."				
RL	J. Bacteriol. 171:6600-6609(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12 / MG1655;				
EX	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RA	Mau B., Shao Y.;				
RT	"The complete genome sequence of Escherichia coli K-12."				
RL	Science 277:1453-1474(1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12;				
EX	MEDLINE=97061202; PubMed=8905232;				
RA	Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,				
RA	Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,				
RA	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,				
RA	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,				
RA	Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yanamoto Y.,				
RA	Yano M., Horiuchi T.;				
RT	"A 718-kb DNA sequence of the Escherichia coli K-12 genome				
RT	corresponding to the 12.7-28.0 min region on the linkage map."				
RL	DNA Res. 3:137-155(1996).				
RN	[4]				
RP	DOMAINS.				
EX	MEDLINE=91296736; PubMed=2068069;				
RA	Levensgood S.K., Beyer W.F. Jr., Webster R.E.;				
RT	"TOLA: a membrane protein involved in colicin uptake contains an				
RT	extended helical region."				
RL	Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).				
RN	[5]				
RP	INTERACTION WITH PORINS.				
EX	MEDLINE=97133271; PubMed=8978668;				
RA	Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,				

34	111	30.3	218	1	H101_CHICK	P08284 gallus gall
35	110.5	30.2	225	1	H15_HUMAN	P16401 homo sapien
36	110.5	30.2	771	1	CALD_CHICK	P12957 gallus gall
37	110	30.1	233	1	H11_GLYBA	P40263 glyptotendi
38	109.5	29.9	194	1	H1_SALTR	P02254 salmo trutt
39	109.5	29.9	209	1	H1A_XENLA	P06892 xenopus lae
40	109	29.8	232	1	H1A_CHITE	P40275 chironomus
41	108	29.5	833	1	IF2_PASMU	P57873 pasteurella
42	107.5	29.4	218	1	H11R_CHICK	P08288 gallus gall
43	106.5	29.1	90	1	H162_TRYCR	P40274 trypanosoma
44	105.5	28.8	235	1	H1E_CHIPA	P40262 chironomus
45	105	28.7	219	1	H1B_XENLA	P06893 xenopus lae

## ALIGNMENTS













```

RX MEDLINE=80156931; PubMed=6767609;
RA Strickland W.N.; Strickland M., de Groot P.C., von Holt C.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides.";
RL Eur. J. Biochem. 104:559-566(1990).
RN [2]
RN SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N.; Strickland M., Brandt W.F., von Holt C., Lehmann A.
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure.";
RL Eur. J. Biochem. 104:567-578(1990).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC -1- NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR: A02586; HSURIP.
DR HSP; P02259; LHST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist N.
DR Pfam; PF00538; linker histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
DR Sperm.
FW VARIANT 144 144 K -> R.
FT SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;
SQ
Query Match 33.6%; Score 123; DB 1; Length 248;
Best Local Similarity 48.7%; Pred. No. 0.0071;
Matches 37; Conservative 9; Mismatches 28; Indels 2; Gaps
200

QY 1 AKYAKKEKAYAKAEKAAK-KAEAKYKAAEKAKKAKAKAEKYYA-KAAAEKKEKYYAAAE 58
Db 145 AKAAKKKALAKKAAAKKAAAKKAAKAAKAKKPKKAAKPKKAAKPKKAAKPKKSP 200
QY 59 AKYKAEAAKAAKAEAA 74
Db 205 KKKKAKRSPKKAKAA 220

RESULT 14
RS6_AEDAE STANDARD; PRT; 346 AA.
ID RS6_AEDAE
AC Q3U761;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S6.
GN RPS6.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Aedes.
OX NCBI_TaxID=7159;
RN [1]
RN SEQUENCE FROM N.A.
RA Gavino V.H., Fallon A.M.;
RL "Aedes mosquitoes ribosomal protein S6 cDNA.";
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH
CC AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR
CC CLASSES OF MRNA (By similarity).
CC -1- FTM: Ribosomal protein S6 is the major substrate of protein
CC kinases in eukaryote ribosomes (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF154067; AAF04790.1; -;  
DR InterPro; IPR001377; Ribosomal\_S6E.  
DR Pfam; PF01092; Ribosomal\_S6e; 1.  
DR ProDom; PD003460; Ribosomal\_S6E; 1.  
DR PROSITE; PS00578; RIBOSOMAL\_S6E; 1.  
KW Ribosomal protein; Phosphorylation.  
SQ SEQUENCE 346 AA; 39365 MW; 599CF7B22BD8DFC CRC64;

Query Match 33.6%; Score 123; DB 1; Length 346;  
Best Local Similarity 45.5%; Pred. No. 0.0092;  
Matches 35; Conservative 9; Mismatches 25; Indels 8; Gaps 2;

Qy 6 KKEKAYAKKAEEKAAK-EEKAYKAAAKKAKAEAKKAYKAAKAEKKEYAA-----A 57

Db 246 EKEKAAKAAKAAKAAKAEKKEYKVTAAKADAKAKTAEKPKAEKSETGKKTAGDKK 305

Qy 58 EAKYKAEAAKAAKEAA 74

Db 306 EKKVEKKAAPAVAKKEA 322

## RESULT 15

ZUOI YEAST  
ID ZUOI YEAST STANDARD; PRT; 433 AA.  
AC P32527;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Zuoitin.  
GN ZUOI OR YGR285C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RC STRAIN=20B-12;  
RX MEDLINE=93010971; PubMed=1396572;  
RA Zhang S., Lockshin C., Herbert A., Winter E., Rich A.;  
RT "Zuoitin, a putative Z-DNA binding protein in Saccharomyces  
RT cerevisiae.";  
RL EMBO J. 11:3787-3796(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY1679;  
RX MEDLINE=37245295; PubMed=9090054;  
RA Volckaert G., Voet M., Robben J.;  
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the  
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying  
RT the MALL locus reveals 15 complete open reading frames, including  
RT ZUOI, BGL2 and BIO2 genes and an ABC transporter gene.";  
RL Yeast 13:251-259(1997).  
CC -1- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN  
CC CHROMOSOME ORGANIZATION.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X63612; CAA45156.1; -;  
DR EMBL; Z73070; CAA97317.1; -;  
DR PIR; S25194; S25194.  
DR HSSP; P08622; 1BQZ.  
DR SGD; S0003517; ZUOI.  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ; 1.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS00636; DNAJ\_1; 1.  
DR PROSITE; PS50076; DNAJ\_2; 1.  
KW Chaperone; DNA-binding; Nuclear protein.  
FT DOMAIN 98 170 J-DOMAIN.  
FT DOMAIN 306 357 ALA/LYS-RICH.  
SQ SEQUENCE 433 AA; 49019 MW; 0AA76BC11D3C7DAB CRC64;

Query Match 33.6%; Score 123; DB 1; Length 433;

Best Local Similarity 46.2%; Pred. No. 0.011;  
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

Qy 2 KKYAKKEKAYAKKAEKAAKAEKAYKAAAEKKAKEAKKYAKA-AKAEK---KEYAAA 57

Db 290 KEEKKEKERRKWEREAGARAEAEAKAEAEAKAEAEAKAEAEAKANASAKADKKKAEAKA 349

Qy 58 EAKYKAEAAKAAKEAAAY 75

Db 350 AKKKCKRAIRNSAKEADY 367

Search completed: March 10, 2003, 12:17:10

Job time : 18.5443 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 40.7743 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989A-5

Perfect score: 366

Sequence: 1 AKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 77

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mmc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_todent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	40.4	394	16 Q8X965	Q8X965 escherichia
2	148	40.4	407	16 Q8Z0P6	Q8Z0P6 salmonella
3	146	39.9	376	16 Q8Z8C1	Q8Z8C1 salmonella
4	145.5	39.8	395	2 Q937K4	Q937K4 erwinia chr
5	142.5	38.9	372	2 Q9WXX1	Q9WXX1 pseudomonas
6	140	38.3	389	16 Q9CM70	Q9CM70 pasteurella
7	138	37.7	232	10 Q39576	Q39576 chlamydomon
8	138	37.7	1701	5 Q61164	Q61164 plasmodium
9	135.5	37.0	1341	16 Q98KG7	Q98KG7 rhizobium 1
10	134	36.6	1128	5 Q26947	Q26947 trypanosoma
11	130.5	35.7	388	16 Q8ZGZ2	Q8ZGZ2 yerlinia pe
12	128	35.0	200	16 Q8XVW7	Q8XVW7 talstonia s
13	125.5	34.3	212	3 Q93946	Q93946 candida alb
14	124.5	34.0	1866	5 Q8T5C8	Q8T5C8 plasmodium
15	122.5	33.5	311	12 Q84528	Q84528 parametium
16	122	33.3	243	16 Q92AC7	Q92AC7 listeria in

17	122	33.3	755	10 Q9FP71	Q9FP71 oryza sativ
18	121.5	32.9	239	16 Q8Y5W4	Q8Y5W4 listeria mo
19	120.5	32.9	284	10 Q65794	Q65794 triticum ae
20	119.5	32.7	191	5 Q46142	Q46142 mytilus edu
21	118.5	32.4	191	5 Q46141	Q46141 mytilus edu
22	118.5	32.4	191	5 Q46143	Q46143 mytilus edu
23	118.5	32.4	191	5 Q46362	Q46362 mytilus edu
24	118.5	32.4	233	5 Q18319	Q18319 chironomus
25	118.5	32.4	275	5 Q01395	Q01395 drosophila
26	118	32.2	969	5 Q9NDI9	Q9NDI9 plasmodium
27	117.5	32.1	265	10 Q39598	Q39598 chlamydomon
28	117.5	32.1	288	10 Q65795	Q65795 triticum ae
29	117	32.0	467	4 Q9H9F1	Q9H9F1 homo sapien
30	116.5	31.8	1671	5 Q8T5C9	Q8T5C9 plasmodium
31	116	31.7	224	13 Q90ZD7	Q90ZD7 bufo bufo g
32	116	31.7	275	10 Q9XHL9	Q9XHL9 triticum ae
33	115.5	31.6	98	5 Q8WQ44	Q8WQ44 leishmania
34	115.5	31.6	111	5 Q8T9R3	Q8T9R3 leishmania
35	115.5	31.6	191	5 Q46140	Q46140 mytilus edu
36	115.5	31.6	243	5 Q23784	Q23784 chironomus
37	115.5	31.6	485	10 Q8RXD0	Q8RXD0 arabidopsis
38	115.5	31.6	924	10 Q9SU08	Q9SU08 arabidopsis
39	114	31.1	68	5 Q26907	Q26907 trypanosoma
40	114	31.1	198	4 Q9H8H4	Q9H8H4 homo sapien
41	114	31.1	237	10 Q9SWU2	Q9SWU2 triticum ae
42	114	31.1	581	16 Q9RU45	Q9RU45 deinococcus
43	114	31.1	607	5 Q9V6S7	Q9V6S7 drosophila
44	113.5	31.0	845	5 Q9Y1P8	Q9Y1P8 plasmodium
45	113	30.9	236	10 Q9SWU3	Q9SWU3 triticum ae

## ALIGNMENTS

### RESULT 1

Q8X965 PRELIMINARY; PRT; 394 AA.

ID Q8X965

AC Q8X965;

DT 01-MAR-2002 (TREMREL. 20, Created)

DT 01-MAR-2002 (TREMREL. 20, Last sequence update)

DT 01-JUN-2002 (TREMREL. 21, Last annotation update)

DE Membrane spanning protein, required for outer membrane integrity

DE (Membrane spanning protein Tola).

GN TOLA OR Z0907 OR ECS0774.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI\_taxid=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Postaf G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"

RN Nature 409:529-533(2001).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.;"

RL DNA Res. 8:11-22(2001).

DR EMBL; AP005252; AAG55075.1; -.

DR EMBL; AP002553; BAB34197.1; -.

DR InterPro; IPR00104; Antifreeze\_1.  
KW PRINTS; PR00308; ANTIFREEZEI.  
KW Complete proteome.  
SQ SEQUENCE 394 AA; 40517 MW; 5B58D8E230BDE28 CRC64;

Query Match 40.4%; Score 148; DB 16; Length 394;  
Best Local Similarity 55.1%; Pred. No. 0.00017;  
Matches 49; Conservative 7; Mismatches 21; Indels 12; Gaps 4;

QY 1 AKKYAKKEKAYA-----KKAEEAAKAEAKAYKAAEA--KKAAEKAYKAA-----KA 49  
DB 138 AKAADADAKAEDDDKAAAEAKAAGAADAKKAAAEAAKAAAEAKAAAEAKAAALKKKA 197  
QY 50 EKKEYAAAAEKYKAAEAACAAA-KEAAYEA 77  
DB 198 EAAEAAAAEARKKAAAEKAAADKKAAEA 226

RESULT 2  
QBZQT6 PRELIMINARY; PRT; 407 AA.

ID QBZQT6 AC  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE TOL protein, membrane spanning protein.  
GN TOLA OR STM0747.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dente M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";  
RL Nature 413:852-856(2001).  
DR EMBL; AB008730; AAL19691.1; -;  
DR InterPro; IPR00104; Antifreeze\_1.  
DR PRINTS; PR00308; ANTIFREEZEI.  
KW Complete proteome.  
SQ SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;

Query Match 40.4%; Score 148; DB 16; Length 407;  
Best Local Similarity 54.4%; Pred. No. 0.00017;  
Matches 49; Conservative 8; Mismatches 13; Indels 20; Gaps 5;

QY 1 AKKYAKKEKAYKAAEKAAKAEAKAYK-AAEAKKAAEKAYKAAEKAAEKAAEKK----EY 54  
DB 189 AKKKAEEAEA-AKAAAEAKKAAEAEEAAKAAAEAKKKADEAAK--AAAEAKKADAAAK 244

QY 55 AAAEAKYKAEAA-----KAAAKEAA 74  
DB 245 AAAEAKKADAAKAAADAKKAAAEAKAA 274

RESULT 3  
QBZC1 PRELIMINARY; PRT; 376 AA.

ID QBZC1 AC  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Tola protein.  
GN STY0793.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;  
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
DR EMBL; AL627288; CAD05209.1; -;  
DR InterPro; IPR000104; Antifreeze\_1.  
DR PRINTS; PR00308; ANTIFREEZEI.  
DR PRO1574; TUBBYPROTEIN.  
KW Complete proteome.  
SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;

Query Match 39.9%; Score 146; DB 16; Length 376;  
Best Local Similarity 54.4%; Pred. No. 0.00023;  
Matches 49; Conservative 7; Mismatches 14; Indels 20; Gaps 5;

QY 1 AKKYAKKEKAYKAAEKAAEKAYK-AAEAKKAAEKAYKAAEKAAEKK----EY 54  
DB 158 AKKKAEEAEA--AKAAADAKKAAEAFAKAAAEAKKGAEEAAK--AAADAKKADAEAAK 213

QY 55 AAAEAKYKAEAA-----KAAAKEAA 74  
DB 214 AAAEAKKADAAKAAKAAADAKKAAAEAKAA 243

RESULT 4  
QB37K4 PRELIMINARY; PRT; 395 AA.

ID QB37K4 AC  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Tola protein.  
GN TOLA.  
OS Erwinia chrysanthemi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Pectobacterium.  
OX NCBI\_TaxID=556;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3937;  
RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;  
RA "Characterization of the Erwinia chrysanthemi tol-pal genes.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ297885; CAC82708.1; -;  
SQ SEQUENCE 395 AA; 41601 MW; 3C0C1DC12E181013 CRC64;

Query Match 39.8%; Score 145.5; DB 2; Length 395;  
Best Local Similarity 61.0%; Pred. No. 0.00026;  
Matches 47; Conservative 9; Mismatches 14; Indels 7; Gaps 5;

QY 1 AKKYAKKE-KAYAKKAEKAAKKA--EAKYKAAEKAAEKAAEKAAEKAAEKYAAA 57  
DB 182 AKKKAEEAAKTAAAEAAAEKAAEKAAEBEAKAAADAKQAEEAK--AKAAEA-AKEKAA 238

QY 58 EAKYKAAEAATAAEKAA 74  
DB 239 DAACKAEAA-AAAKKAA 254

RESULT 5



RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RL	"Complete genomic sequence of Pasteurella multocida Pm70.";
RT	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR	EMBL; AB006136; AAK03052.1; -.
DR	HSSP; PI9934; ITOL.
DR	InterPro; IPR005533; Tropomyosin.
DR	PRINTS; PR00194; TROPOMYOSIN.
KW	Complete proteome.
SQ	SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9B94B CRC64;
 Query Match                  38.3%; Score 140; DB 16; Length 389; Best Local Similarity      50.5%; Pred.No. 0.0007; Matches 46; Conservative 10; Mismatches 21; Indels 14; Gaps	
QY	1 AKYYA-----KKEKAYAKAEAAKAKE--AKAYKAAEA---KKGAKEAEAKKYAKAAKAE- 50     : :      ::         ::     ::     ::     ::
Db	167 AKRLAALAKQAEEAKKAAEAEAKRKAEEKAKAEAEAKVEKAKAEAEAKVVAEKAKAEA 226     : :      ::         ::     ::     ::     ::
QY	51 -----KKEYAAAAEAKYKAEAAKAAKAAEAYEA 77                                ::
Db	227 EAKAKAEAKAEAKAEAKAEAKADAEAAQRKA 257                                ::
 RESULT 7 Q39576 PRELIMINARY; PRT; 232 AA. ID C Q39576; AC Q39576; DT 01-NOV-1996 (TrEMBLrel_01, Created) DT 01-NOV-1996 (TrEMBLrel_01, Last sequence update) DT 01-MAR-2002 (TrEMBLrel_20, Last annotation update) DE Histone H1. GN CH1. OS Chlamydomonas reinhardtii. OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; OC Chlamydomonadaceae; Chlamydomonas. OX NCBI_TaxID=3055; RN [1] RP SEQUENCE FROM N.A. RX MEDLINE=96120862; PubMed=8590479; RA Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R. RT "The organization structure and regulatory elements of Chlamydomonas histone genes reveal features linking plant and animal genes."; RL Curr. Genet. 28:333-345(1995). DR EMBL; UI6726; AAA98452.1; -. DR HSSP; P02259; IHST. DR InterPro; IPR001386; Histone_H1/H5. DR InterPro; IPR003216; Linkerhist_N. DR Pfam; PF00538; linker_histone; 1. DR ProDom; PD000373; Linkerhist_N; 1. DR SMART; SM00526; H15; 1. SQ SEQUENCE 232 AA; 24693 MW; 2D006AE44A8FA037 CRC64;	
 Query Match                  37.7%; Score 138; DB 10; Length 232; Best Local Similarity      49.4%; Pred.No. 0.00061; Matches 41; Conservative 4; Mismatches 28; Indels 10; Gaps	
QY	1 AKYKAKEKAYAKAEAAKAAKAE-----KAYKAEAEKKKAEAAKAYAKAAKAEK 51 
Db	103 AKKAATPKAAAPKKGAVKTKTPAPKKEGKKPKSAKAEKPKKEGKKKAAKPAKAEK 162 
QY	52 KEYAAAEEAKYKAEAAKAAKAAEAA 74 
Db	163 KPAAKPAK-KTPTKKAAPKA 184 
 RESULT 8 O61164 PRELIMINARY; PRT; 1701 AA. ID O61164; AC O61164; DT 01-AUG-1998 (TrEMBLrel_07, Created) DT 01-AUG-1998 (TrEMBLrel_07, Last sequence update) DT 01-OCT-2000 (TrEMBLrel_15, Last annotation update) DE Erythrocyte binding protein.	



Db 155 KAKEQKQETAAQAQKADKIVKAQAEKQKAEAKKEAAVAAAKKQ-ADADAKKA 213  
QY 63 ABAKAAKAAKEAA 74  
Db 214 VEVAEKAAADAA 225  
RESULT 12  
Q8XV7 PRELIMINARY; PRT; 200 AA.  
AC Q8XV7;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Probable histone H1 protein.  
GN RSC2793 OR RS00453.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,  
Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646071; CAD16500.1; -;  
DR InterPro; IPR00104; Antifreeze\_1.  
DR PRINTS; PRO0308; trscalc.  
DR PRINTS; PRO1503; ANTIFREEZE1.  
KW Complete proteome.  
SQ SEQUENCE 200 AA; 19279 MW; D3831B590510272D CRC64;  
Query Match 35.0%; Score 128; DB 16; Length 200;  
Best Local Similarity 50.0%; Pred. No. 0.0032;  
Matches 43; Conservative 3; Mismatches 28; Indels 12; Gaps 3;  
QY 1 AKKYAKKEKAYAKKAEKAAKAEAK-----AVKAAEAKKA--KAEKVKAAKAAE 50  
Db 19 AAKKAPAKKAVAKKAPVAKKAPAKKVAKKVAAKKAAPAKKAAVKKVAKKAAKAA 78  
QY 51 KKEYAA--AAKYYKAEKAAKAAKAAEAA 74  
Db 79 VKKVAKKAPAAKAAKAAVKKVAKKAA 104  
RESULT 13  
O93946 PRELIMINARY; PRT; 212 AA.  
AC O93946;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE CTA2p (Fragment).  
GN CTA2.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kaiser B., Kunkel W., Saluz H.P., Munder T.;  
RT "Identification of Candida albicans protein domains with  
transcriptional activating properties in Saccharomyces cerevisiae.";  
RL submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ006637; CAA07165.1; -;  
FT NON TER 1  
SQ SEQUENCE 212 AA; 24231 MW; 10C2122E9554A387 CRC64;  
Query Match 34.3%; Score 125.5; DB 3; Length 212;  
Best Local Similarity 47.9%; Pred. No. 0.0054;  
Matches 34; Conservative 9; Mismatches 27; Indels 1; Gaps 1;  
QY 7 KEKAYAKKAEKAAKAEKAYKAAEAKKAKAEAKKAEKAAKAEKAAEAKYKAAEAA 66  
Db 51 KEEEAARKKEEAKKKEEAKKAEAKKKEEAKKAEAKKAEAKKAEAK-KVEEAARKAEAA 109  
QY 67 KAAAEKAAAYEA 77  
Db 110 KKAEEAEARKKA 120  
RESULT 14  
Q8T5C8 PRELIMINARY; PRT; 1866 AA.  
AC Q8T5C8;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Maeb1.  
GN Plasmodium vivax.  
OS Plasmodium vivax.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5855;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SAIVADOR;  
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;  
RT "Evolutionary relationships of conserved cysteine-rich motifs in  
RT adhesive molecules of malaria parasites.";  
RL Mol. Biol. Evol. 0:0-0(2002).  
DR EMBL; AY042083; AAL10508.1; -;  
SQ SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;  
Query Match 34.0%; Score 124.5; DB 5; Length 1866;  
Best Local Similarity 50.0%; Pred. No. 0.054;  
Matches 42; Conservative 10; Mismatches 23; Indels 9; Gaps 4;  
QY 1 AKKYAKKEKAYAKKAEKAAKAEKAAE--KAYKAAEAKKKKAEAKKAYKAAEAKKAEK 54  
Db 1274 AKKAEAEARKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAAKAEAA 1333  
QY 55 A-AEAKYKAEAAKAAKAAEAAAYEA 77  
Db 1334 AKKAEAEARKAEAAKKA--EAARKA 1355  
RESULT 15  
Q84528 PRELIMINARY; PRT; 311 AA.  
AC Q84528;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE A208R protein.  
GN A208R.  
OS Paramesidium bursaria chlorella virus 1 (PBCV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.  
OX NCBI\_TaxID=10506;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96187795; PubMed=8614977;  
RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;  
RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map  
RT positions 88 to 182.";  
RL Virology 216:102-123(1996).  
RN [2]  
RP SEQUENCE FROM N.A.

```

RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., King W.,
RA Lisee A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homosperrmidine
RT synthase.";
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PBCV-1.";
RL Virology 276:27-36(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42580; AAC96576.1; -;
DR InterPro; IPR02048; EF-hand.
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.
SQ SEQUENCE 311 AA; 35390 MW; 97B0E9E82AFEF88A CRC64;

Query Match 33.5%; Score 122.5; DB 12; Length 311;
Best Local Similarity 48.6%; Pred. No. 0.013;
Matches 36; Conservative 9; Mismatches 26; Indels 3; Gaps 2;

Qy 1 AKKYAKKEKAYAKK-AEKAAYKAAKAYKAAEAK--KKAAEAKKYAKAAKAEKYYAA 57
Db 66 AEKAAEKERVKAEEAAREAEKERVKAEEKAEKAEKAEKAEKAEKAEKAEKAEKAEK 125
Qy 58 EAKYKAEAAKAAK 71
Db 126 EKAPAREAKAAK 139

```

Search completed: March 10, 2003, 12:25:57  
Job time : 41.7743 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 40.1245 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989A-5

Perfect score: 366

Sequence: 1 AKKYAKKEKAYAKAEKAAK.....EAKYKAEAKAAKAEAYEA 77

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	366	100.0	77	21	AAV82575
2	351.5	96.0	86	21	AAV82576
3	289	79.0	109	21	AAV82577
4	282.5	77.2	66	21	AAV82574
5	245.5	67.1	56	21	AAV82573
6	144.5	39.5	100	21	AAV98499
7	144.5	39.5	100	21	AAV59044
8	144.5	39.5	100	22	AAU04289
9	144.5	39.5	100	22	AAU45852
10	142.5	38.9	154	11	AAV06445
					Copolymer molecule
					Copolymer molecule
					Copolymer molecule
					Copolymer molecule
					Copolymer molecule
					Peptide #10 used i
					Amino acid polymer
					Poly-lys-Ala used
					Nucleic acid trans
					Recombinant copoly

11	140	38.3	106	11	AAV06446	Recombinant copoly
12	134	36.6	643	16	AAV84568	Trypanosoma cruzi
13	131	35.8	45	21	AAV82572	Copolymer molecule
14	130	35.5	205	16	AAV820575	Mycobacterium bovi
15	130	35.5	472	16	AAV84569	Trypanosoma cruzi
16	130	35.5	564	16	AAV84565	Trypanosoma cruzi
17	129.5	35.4	214	20	AAV34055	M. tuberculosis hi
18	129.5	35.4	214	20	AAV34055	M. tuberculosis hi
19	123	33.6	433	18	AAV30256	Zuotin. Saccharom
20	123	33.6	433	20	AAV32954	Human zuotin prote
21	123	33.6	433	22	AAV70739	S cerevisiae apopt
22	122.5	33.5	140	13	AAV25206	Synthetic helical
23	121.5	33.2	239	23	ABB49123	Listeria monocytog
24	121	33.1	223	20	AAV14928	Amino acid sequenc
25	117	32.0	279	22	AAU03592	Human DNA modifca
26	117	32.0	467	22	AAV94309	Human protein sequ
27	116	31.7	427	22	AAV70868	C albicans apoptos
28	114.5	31.3	219	21	AAV76981	Quail H1 histone p
29	114	31.1	198	22	AAV95499	Human protein sequ
30	114	31.1	607	22	ABB62173	Drosophila melanog
31	114	31.1	1507	21	AAV24128	Plasmodium yoelii
32	112.5	30.7	146	22	AAV25508	Human protein sequ
33	112	30.6	222	18	AAV29476	Human histone H1 i
34	112	30.6	222	18	AAV34033	Human histone H1 i
35	112	30.6	222	21	AAV57331	Human histone H1 i
36	110.5	30.2	158	20	AAV34068	Histone H1 isoform
37	110.5	30.2	158	21	AAV57366	Human histone H1.5
38	110.5	30.2	226	20	AAV34060	Human histone H1 i
39	110.5	30.2	226	21	AAV57358	Human histone H1 i
40	110	30.1	334	22	ABG28693	Novel human diagno
41	108	29.5	130	23	AAU09945	Peptide fragment o
42	108	29.5	130	23	AAE13237	Human linker histo
43	107.5	29.4	262	17	AAV06913	T. cruzi L19E homo
44	107.5	29.4	262	20	AAV32848	TcE protein sequen
45	107.5	29.4	263	20	AAV23306	TcE antigen of Trp

#### ALIGNMENTS

#### RESULT 1

AAV82575  
ID AAV82575 standard; peptide; 77 AA.

XX AAV82575;

XX AC

DT 28-JUL-2000 (first entry)

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
XX antidiabetic; chymotrypsin; haemostatic; antipsoriatic; dermatological;  
XX antianemic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
XX pemphigus vulgaris; systemic lupus erythematosus.

OS Unidentified.

XX WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.



inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
 diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 pemphigus vulgaris; systemic lupus erythematosus.  
 Unidentified.  
 WO200018794-A1.  
 06-APR-2000.  
 24-SEP-1999; 99WO-US22402.  
 25-SEP-1998; 98US-0101693.  
 (YEDA ) YEDA RES & DEV CO LTD.  
 (TEVA-) TEVA PHARM USA INC.  
 Gad A, Lis D;  
 WPI; 2000-317499/27.  
 Copolymer 1 related polypeptides used as molecular weight markers for  
 glatiramer acetate and for treatment and prevention of immune diseases  
 Claim 10; Page 14; 72pp; English.  
 AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 weight TV-marker polypeptides from the present invention. The present  
 invention describes polypeptides (I) for determining the molecular  
 weight of a copolymer (CP), which has an identified molecular weight  
 and an amino acid composition corresponding to the copolymer. The  
 polypeptides of the invention are used as molecular weight markers for  
 glatiramer acetate related tetrapolymers. The polypeptides may also be  
 used for treating and preventing immune diseases in a mammal. Autoimmune  
 diseases which may be treated include either cell-mediated or  
 antibody-mediated diseases. Such diseases include arthritic conditions,  
 demyelinating diseases and inflammatory conditions, e.g. multiple  
 sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
 purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 lupus erythematosus. Mediated-mediated diseases which can be treated  
 include host-versus-graft disease, graft-versus-host disease, and  
 delayed-type hypersensitivity. The polypeptides of the invention have  
 defined molecular weights and physical properties which are analogous to  
 glatiramer acetate molecules, which makes them ideal for use as  
 molecular weight markers.  
 Sequence 109 AA;  
 Query Match 79.0%; Score 289; DB 21; Length 109;  
 Best Local Similarity 67.0%; Pred. No. 6.5e-20;  
 Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4;  
 QY 1 AKKYAKK-EKAYAKA-----EKAAKKAEKAYAKAEAKKA----- 36  
 DB 1 AKKYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAE 60  
 QY 37 -KAEAKKYAKKAEKAYAKAEAK-----YKAEAKKAEKAEKAE 77  
 DB 61 YKAEAKKYAKKAEKAYAKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 109  
 RESULT 4  
 AAY82574  
 ID AAY82574 standard; peptide; 66 AA.  
 XX  
 AC AAY82574;  
 XX

DT 28-JUL-2000 (first entry)  
 XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.  
 DE  
 XX  
 KW Copolymer; molecular weight marker; TV-marker; immune disease;  
 KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
 KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
 KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
 KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
 KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
 KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
 KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
 KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
 KW pemphigus vulgaris; systemic lupus erythematosus.  
 OS Unidentified.  
 OS WO200018794-A1.  
 PN 06-APR-2000.  
 XX 24-SEP-1999; 99WO-US22402.  
 XX 25-SEP-1998; 98US-0101693.  
 PR (YEDA ) YEDA RES & DEV CO LTD.  
 XX (TEVA-) TEVA PHARM USA INC.  
 PA  
 PA Gad A, Lis D;  
 XX  
 PI WPI; 2000-317499/27.  
 XX  
 DR Copolymer 1 related polypeptides used as molecular weight markers for  
 XX glatiramer acetate and for treatment and prevention of immune diseases  
 PT Claim 10; Page 14; 72pp; English.  
 XX  
 PS AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
 CC weight TV-marker polypeptides from the present invention. The present  
 CC invention describes polypeptides (I) for determining the molecular  
 CC weight of a copolymer (CP), which has an identified molecular weight  
 CC and an amino acid composition corresponding to the copolymer. The  
 CC polypeptides of the invention are used as molecular weight markers for  
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
 CC used for treating and preventing immune diseases in a mammal. Autoimmune  
 CC diseases which may be treated include either cell-mediated or  
 CC antibody-mediated diseases. Such diseases include arthritic conditions,  
 CC demyelinating diseases and inflammatory conditions, e.g. multiple  
 CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
 CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
 CC lupus erythematosus. Mediated-mediated diseases which can be treated  
 CC include host-versus-graft disease, graft-versus-host disease, and  
 CC delayed-type hypersensitivity. The polypeptides of the invention have  
 CC defined molecular weights and physical properties which are analogous to  
 CC glatiramer acetate molecules, which makes them ideal for use as  
 CC molecular weight markers.  
 XX Sequence 66 AA;  
 SQ Query Match 77.2%; Score 282.5; DB 21; Length 66;  
 Best Local Similarity 84.4%; Pred. No. 1.5e-19;  
 Matches 65; Conservative 0; Mismatches 1; Indels 11; Gaps 2;  
 QY 1 AKKYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAE 60  
 DB 1 AKKYAKKAEKAYAK-----AKKAEK-----AKKAEKAEKAYAKKAEKAEKAEKAE 49  
 QY 61 YKAEAKKAEKAEKAEKAEKAE 77  
 ||||||||||||||||

Db 50 YKAEAAKAAKEAAAYEA 66  
RESULT 5  
RAY82573  
ID AAY82573 standard; peptide; 56 AA.  
XX  
AC AAY82573;  
XX  
DT 28-JUL-2000 (first entry)  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.  
XX  
KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antinaeemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX  
OS Unidentified.  
XX  
XX WO200018794-A1.  
XX  
XX PN 06-APR-2000.  
XX  
XX PD 24-SEP-1999; 99WO-US22402.  
XX  
XX PF 25-SEP-1998; 98US-0101693.  
XX  
XX PR (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX PA (TEVA-) TEVA PHARM USA INC.  
XX  
XX PI Gad A, Lis D;  
XX  
XX WPI; 2000-317499/27.  
XX  
XX DR Copolymer 1 related polypeptides used as molecular weight markers for  
XX glatiramer acetate and for treatment and prevention of immune diseases  
XX  
XX PS Claim 10; Page 14; 72pp; English.  
XX  
XX CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
XX weight TV-marker polypeptides from the present invention. The present  
XX invention describes polypeptides (I) for determining the molecular  
XX weight of a copolymer (CP), which has an identified molecular weight  
XX and an amino acid composition corresponding to the copolymer. The  
XX polypeptides of the invention are used as molecular weight markers for  
XX glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX used for treating and preventing immune diseases in a mammal. Autoimmune  
XX diseases which may be treated include either cell-mediated or  
XX antibody-mediated diseases. Such diseases include arthritic conditions,  
XX demyelinating diseases and inflammatory conditions, e.g. multiple  
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia  
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX lupus erythematosus. Mediated-mediated diseases which can be treated  
XX include host-versus-graft disease, graft-versus-host disease, and  
XX delayed-type hypersensitivity. The polypeptides of the invention have  
XX defined molecular weights and physical properties which are analogous to  
XX glatiramer acetate molecules, which makes them ideal for use as  
XX molecular weight markers.  
XX  
XX SQ Sequence 56 AA;  
Query Match 67.1%; Score 245.5; DB 21; Length 56;

Best Local Similarity 72.7%; Pred. No. 3.5e-16;  
Matches 56; Conservative 0; Mismatches 0; Indels 21; Gaps 1;  
QY 1 AKKYAKKAYAKKAEKAAKAEKAYKAAEAKKKAKAEAKKYAKAAKAEKKEVAAAEAK 60  
DB 1 AKKYAKKAYAKKAEKAAKAEKAYKAAEAKKKAKAEAKKYAKAAKAEKKEVAAAEAK 39  
QY 61 YKAEAAKAAKAAKEAAAYEA 77  
DB 40 YKAEAAKAAKAAKEAAAYEA 56  
RESULT 6  
AAY98499  
ID AAY98499 standard; Peptide; 100 AA.  
XX  
AC AAY98499;  
XX  
DT 31-JUL-2000 (first entry)  
XX  
DE Peptide #10 used in nucleic acid transporter system.  
XX  
KW Transporter system; nucleic acid delivery; gene therapy; cancer;  
KW carcinogenesis; cardiovascular disease; infection.  
XX  
OS Synthetic.  
XX  
XX PN US6033884-A.  
XX  
XX PD 07-MAR-2000.  
XX  
XX PF 14-DEC-1993; 93US-0167641.  
XX  
XX PR 20-MAR-1992; 92US-0855389.  
XX  
XX PR 19-MAR-1993; 93WO-US02725.  
XX  
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
XX PI Gottchalk S, Sparrow J, Cristiano RJ, Woo SLC, Smith LC;  
XX  
XX WPI; 2000-281993/24.  
XX  
XX PT System for transporting nucleic acid into cells, useful e.g. in gene  
XX therapy and for generating transgenic animals, comprises binding agent  
XX linked to nucleic acid, surface ligand and lytic agent -  
XX  
XX PS Disclosure; Column 125-128; 108pp; English.  
XX  
XX CC The present invention relates to a transporter system for delivering  
XX nucleic acid to a cell. The system comprises a nucleic acid binding  
XX complex, consisting of a binding molecule bonded non-covalently to the  
XX nucleic acid, and covalently to a surface ligand, and a lytic agent. The  
XX binding molecule is spermine or a spermidine derivative. Nucleotide  
XX sequences AAA36633-A36652 and peptide sequences AAY98456-Y98500 are used  
XX in the construction of the transporter system of the invention. The  
XX transporter system is used in gene therapy, particularly to deliver  
XX nucleic acids to hepatocytes, muscle cells or bone forming cells, e.g.  
XX for treating cardiovascular disease, cancer, and infection. The  
XX transporter systems are also used to create transgenic animals (as models  
XX for human carcinogenesis or disease or for drug testing). Other uses  
XX include transfecting cells to produce proteins, or transfecting cells in  
XX vitro to study the function of the nucleic acid. The use of a surface  
XX ligand allows specific targeting of selected cells and tissues. The lytic  
XX agent provides for release of the nucleic acid into the cellular  
XX interior, from endosomes, without requiring endosomal or lysosomal  
XX degradation.  
XX  
XX SQ Sequence 100 AA;  
Query Match 39.5%; Score 144.5; DB 21; Length 100;  
Best Local Similarity 58.4%; Pred. No. 1.5e-06;  
Matches 45; Conservative 6; Mismatches 23; Indels 3; Gaps 3;









Db 547 TKVAEAEKQKAAEAA 561

## RESULT 13

RAY82572  
ID AAY82572 standard; peptide; 45 AA.

XX AC AAY82572;

XX DT 28-JUL-2000 (first entry)

XX DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:2.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;  
XX KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
XX KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
XX KW antidiabetic; thyronimetic; haemostatic; antipsoriatic; dermatological;  
XX KW antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
XX KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
XX KW pemphigus vulgaris; systemic lupus erythematosus.

XX OS Unidentified.

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PA (TEVA-) TEVA PHARM USA INC.

XX PI Gad A, Lis D;

XX WPI; 2000-317499/27.

XX FT Copolymer 1 related polypeptides used as molecular weight markers for  
XX FT glatiramer acetate and for treatment and prevention of immune diseases

XX PS Claim 10; Page 14; 72pp; English.

XX CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
XX CC weight TV-marker polypeptides from the present invention. The present  
XX CC invention describes polypeptides (I) for determining the molecular  
XX CC weight of a copolymer (CP), which has an identified molecular weight  
XX CC and an amino acid composition corresponding to the copolymer. The  
XX CC polypeptides of the invention are used as molecular weight markers for  
XX CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
XX CC used for treating and preventing immune diseases in a mammal. Autoimmune  
XX CC diseases which may be treated include either cell-mediated or  
XX CC antibody-mediated diseases. Such diseases include arthritic conditions,  
XX CC demyelinating diseases and inflammatory conditions, e.g. multiple  
XX CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
XX CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
XX CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
XX CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
XX CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
XX CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
XX CC lupus erythematosus. Mediated-mediated diseases which can be treated  
XX CC include host-versus-graft disease, graft-versus-host disease, and  
XX CC delayed-type hypersensitivity. The polypeptides of the invention have  
XX CC defined molecular weights and physical properties which are analogous to  
XX CC glatiramer acetate molecules, which makes them ideal for use as  
XX CC molecular weight markers.

SQ Sequence 45 AA;

## Query Match

35.8%; Score 131; DB 21; Length 45;

Best Local Similarity 50.6%; Pred. No. 1.1e-05;  
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

QY 1 AKKYAKKEKAYAKKAAKKAAYKAAEAKKKAKAEAKKYAKAKAEKKEAYAAAEAK 60  
DB 1 AKKYAKK-----AKAEKA-----KKAYKAAEAKKAAKYE----- 29

QY 61 YKAEAAKAAKAAEAYEA 77

DB 30 -KAAEKAKEAYEA 45

## RESULT 14

AAB20575

ID AAB20575 standard; Protein; 205 AA.

XX AC AAB20575;

XX DT 08-DEC-2000 (first entry)

XX DE Mycobacterium bovis BCG strain MDPI protein SEQ ID NO:2.

XX KW Mycobacterium bovis BCG strain; mycobacterium DNA binding protein 1;  
XX KW MDPI; slow growing acid-fast bacterium protein; immunogenicity;  
XX KW pathogenic acid-fast bacterium; diagnosis; tuberculosis; MAC;  
XX KW Mycobacterium avium intracellular complex; Hansen's disease.

XX OS Mycobacterium bovis.

XX PN WO200044905-A1.

XX PD 03-AUG-2000.

XX PF 28-JAN-2000; 2000WO-JP00455.

XX PR 29-JAN-1999; 99JP-0022588.

XX PA (SAKA ) OTSUKA PHARM CO LTD.

XX PA (YAMA/) YAMADA T.

XX PA (MATS/) MATSUMOTO S.

XX PI Yamada T, Matsumoto S;

XX WPI; 2000-543393/49.

XX DR N-PSDB; AAA88001.

XX FT New Mycobacterium DNA binding protein 1 immunogenic for pathogenic  
XX FT acid-fast bacteria for diagnosis, treatment and prevention of diseases  
XX FT such as tuberculosis

XX PS Claim 1; Fig 2A; 62pp; Japanese.

XX CC The present sequence represents the Mycobacterium bovis BCG strain  
XX CC mycobacterium DNA binding protein 1 (MDPI), which has immunogenicity  
XX CC against pathogenic acid-fast bacteria. The MDPI protein can be used for  
XX CC diagnosis and treatment of diseases involving pathogenic acid-fast  
XX CC bacteria, such as tuberculosis, Mycobacterium avium intracellular  
XX CC complex (MAC) and Hansen's disease.

XX SQ Sequence 205 AA;

Query Match 35.5%; Score 130; DB 21; Length 205;

Best Local Similarity 49.4%; Pred. No. 7.2e-05;  
Matches 42; Conservative 7; Mismatches 26; Indels 10; Gaps 3;

QY 1 AKKYAKKEKAYAKKAAKKA--EAKYKAA-----EAKKAKAEAKKYAKAKAEKK 52  
DB 111 AKKVAK--KAPAKKATKAAKKAATKAPAKKATKAPAKKAVKATKSPAKKVTKAVKKTAV 168

QY 53 EYAAAEAKYKAEAAKAAKEAYEA 77

DB 169 KASVRKAAATKAPAKKAAAKRPATKA 193



THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match		Length	DB	ID	Description
1	366	100.0	77	10	US-09-816-989A-5	Sequence 5, Appl	
2	351.5	96.0	86	10	US-09-816-989A-6	Sequence 6, Appl	
3	289	79.0	109	10	US-09-816-989A-7	Sequence 7, Appl	
4	282.5	77.2	66	10	US-09-816-989A-4	Sequence 4, Appl	
5	245.5	67.1	56	10	US-09-816-989A-3	Sequence 3, Appl	
6	141	38.5	372	9	US-09-820-843A-8	Sequence 8, Appl	
7	131	35.8	45	10	US-09-816-989A-2	Sequence 2, Appl	
8	124.5	34.0	452	9	US-10-184-832-5	Sequence 5, Appl	
9	121	33.1	223	9	US-10-051-643-201	Sequence 201, App	
10	111.5	30.5	356	9	US-09-820-843A-27	Sequence 27, Appl	
11	104.5	28.6	35	10	US-09-816-989A-1	Sequence 1, Appl	
12	104.5	28.6	309	9	US-09-820-843A-24	Sequence 24, Appl	
13	104.5	28.6	617	10	US-09-864-761-36182	Sequence 36182, A	
14	102	27.9	369	9	US-09-820-843A-95	Sequence 95, Appl	
15	99	27.0	102	9	US-09-999-724-90	Sequence 90, Appl	
16	99	27.0	418	9	US-09-999-724-48	Sequence 48, Appl	
17	99	27.0	234	10	US-09-124-280A-3	Sequence 3, Appl	
18	98.5	26.9	892	10	US-09-815-242-13765	Sequence 13765, A	
19	97.5	26.6	890	10	US-09-815-242-10314	Sequence 10314, A	





```

QY      1 AKKYAKKEKAYAKAEAAKKAAEAKYKAAABAKKKAEEAKKYAKAAKAEKKEYAAAAEK 60
        |||||         |||||         |||||         |||||         |||||
Db      1 AKKYAKK-----AKAEEA-----KKAYKAAEAEEKAAKYE-----29

QY      61 YKAAEAKAAAEKAAEAYA 77
        || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     30 -KAAAEKAAAEKAAEAYA 45

RESULT 8
US-10-184-832-5
; Sequence 5, Application US/10184832
; Publication No. US20030022857A1
; GENERAL INFORMATION:
; APPLICANT: Xu et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITL OF INVENTION: TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; FILE REFERENCE: WPI2001-056P1NW
; CURRENT APPLICATION NUMBER: US/10/184,832
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: 60/303,250
; PRIOR FILING DATE: 2001-07-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-184-832-5

Query Match          34.0%; Score 124.5; DB 9; Length 452;
Best Local Similarity 56.0%; Pred. No. 0.00041;
Matches 42; Conservative 7; Mismatches 21; Indels 5; Gaps 4;

QY      1 AKKYAKE-KAYAKAAEAKKAAEAKYKAAEAKKKKAAEAKKYAKA-AKAEKKEYAAAAE 58
        || | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db     94 AEAEAKAEPKAAEAEAPKAAEAEAPKA--EAENEAEAKAAEAKAAEAKAAEAKAE 151
        || | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY      59 AKYKAEA-AKAAAKE 72
        || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     52 AKAEKAAEAKAAEAKE 166

RESULT 9
US-10-051-643-201
; Sequence 201, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITL OF INVENTION: Of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

Query Match          33.1%; Score 121; DB 9; Length 223;
Best Local Similarity 55.6%; Pred. No. 0.0004;
Matches 45; Conservative 4; Mismatches 22; Indels 10; Gaps 5;
```

Qy	1	AKVYKKEKAYAKK--AKKAAPKAEEKAYKAAAKKKAAE-----AKVYKAAKAEKKEY 54
Db	121	AKCAAKKAPKAAPKAAPKAATKA-APAKKATATKKAAPAKKATATKKAAPKAAPKA 179
Qy	55	A--AAEAKYKAEKAAKAAKEA 73
Db	180	ATKAAPAK-KAPAKKAAATKAA 199

```

RESULT 10
US-09-820-843A-27
; Sequence 27, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: tola protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|9656364
; US-09-820-843A-27

```

```

RESULT 11
US-09-816-989A-1
; Sequence 1, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-1
Query Match          28.6%; Score 104.5; DB 10; Length 35;

```

```
Best Local Similarity    65.8%;   Pred. No. 0.0016;
Matches      25; Conservative       4; Mismatches     6; Indels      3; Gaps      1;

QY    40 AKKYAKAAKKKYYAAAEEKYKAEAAKAATAAEEAYEA 77
||||| ||||| : |:::||||| 
Db     1 AKKYAKEKAAKKAY---KKEAKAKAAEAAEKAEAYEA 35

RESULT 12
US-09-820-843A-24
; Sequence 24, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROT
; TITLE OF INVENTION: USEFUL AS ANTI-INPECTIVES
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: polyhydroxyalkanoate synthesis protein Phaf
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|9951352
; US-09-820-843A-24

Query Match          28.6%   Score 104.5; DB 9; Length 309;
Best Local Similarity 48.9%;   Pred. No. 0.015;
Matches      43; Conservative       0; Mismatches    34; Indels    11; Gaps      4;

QY    1 AKKYAK--KEKAYAKAEKAAKAEAK-----AYKAAEAKKKAKAEAKKYAK-AAKAEK 51
Db    154 KPAAKTAAAKPAAPKAAKAAKAAKPAAKTAAKTAAKPAAPKAAKPTAKAAAKPAT 213

QY    52 KEYAAREAK--YKAEAAKAAKEAAEAYEA 77
| | | | | | | | | | | | | | | | | |
Db    214 KPAKAAAKPAAKPAAPKAAKPAAPKAAKPA 241

RESULT 13
US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn. Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
```



THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 13.6456 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-5  
Perfect score: 366  
Sequence: 1 AKYAKKEKAYAKKAERAAK.....EAKYKAEAKAAKAEAYEA 77

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144.5	39.5	100	2	US-08-460-890A-64
2	144.5	39.5	100	3	US-08-167-641C-64
3	144.5	39.5	100	4	US-08-460-971A-64
4	144.5	39.5	100	4	US-08-462-040-64
5	134	36.6	643	2	US-08-216-894-8
6	134	36.6	643	4	US-09-115-746-8
7	130	35.5	472	2	US-08-216-894-10
8	130	35.5	472	4	US-09-115-746-10
9	130	35.5	564	2	US-08-216-894-2
10	130	35.5	564	4	US-09-115-746-2
11	129.5	35.4	214	3	US-09-041-889-27
12	123	33.6	433	1	US-08-346-849-2
13	123	33.6	433	2	US-08-293-284A-2
14	121	33.1	223	4	US-09-095-855-201
15	121	33.1	223	4	US-09-205-426-201
16	119.5	32.7	56	4	US-08-993-008A-6
17	114	31.1	1507	3	US-08-929-329-5
18	112	30.6	222	3	US-09-041-889-3
19	112	30.6	222	3	US-08-837-058-3
20	110.5	30.2	158	3	US-09-041-889-40
21	110.5	30.2	226	3	US-09-041-889-32
22	107.5	29.4	262	1	US-08-403-379A-1
23	107.5	29.4	262	2	US-08-929-414-1
24	107.5	29.4	263	2	US-08-557-309B-51
25	107.5	29.4	263	3	US-08-834-306-51
26	107.5	29.4	263	4	US-08-993-674A-51
27	107.5	29.4	263	4	US-09-256-976-51

28 106.5 29.1 48 4 US-08-993-008A-5 Sequence 5, Appli  
29 102 27.9 218 3 US-09-041-889-4 Sequence 4, Appli  
30 102 27.9 218 3 US-08-837-058-4 Sequence 4, Appli  
31 101.5 27.7 60 1 US-08-346-849-16 Sequence 16, Appl  
32 101.5 27.7 60 2 US-08-293-284A-16 Sequence 16, Appl  
33 99.5 27.2 116 3 US-09-041-889-38 Sequence 38, Appl  
34 99 27.0 100 2 US-08-460-890A-62 Sequence 62, Appl  
35 99 27.0 100 3 US-08-167-641C-62 Sequence 62, Appl  
36 99 27.0 100 4 US-08-460-971A-62 Sequence 62, Appl  
37 99 27.0 100 4 US-08-462-040-62 Sequence 62, Appl  
38 99 27.0 102 4 US-09-101-751A-90 Sequence 90, Appl  
39 99 27.0 218 4 US-09-101-751A-48 Sequence 48, Appl  
40 99 27.0 434 1 US-08-097-830B-3 Sequence 3, Appli  
41 99 27.0 434 2 US-08-456-112B-3 Sequence 2, Appli  
42 98.5 26.9 92 4 US-09-344-529-2 Sequence 54, Appl  
43 98 26.8 219 2 US-08-557-309B-54 Sequence 52, Appl  
44 97 26.5 442 3 US-08-834-306-52 Sequence 52, Appl  
45 97 26.5 442 4 US-08-993-674A-52 Sequence 52, Appl

## ALIGNMENTS

RESULT 1  
US-08-460-890A-64  
; Sequence 64, Application US/08460890A  
; Patent No. 5994109  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,890A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/167,641  
; FILING DATE: December 14, 1993  
; APPLICATION NUMBER: 07/855,389  
; FILING DATE: March 20, 1992  
; APPLICATION NUMBER: PCT/US93/02725  
; FILING DATE: March 19, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 212/066  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELE: 67-3510  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

;  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: "Lys Ala" in positions 3 to 100 may be  
; OTHER INFORMATION: present or absent.  
US-08-460-890A-64

Query Match 39.5%; Score 144.5; DB 2; Length 100;  
Best Local Similarity 58.4%; Pred. No. 2.7e-07;  
Matches 45; Conservative 6; Mismatches 23; Indels 3; Gaps 3;

Qy 1 AKKYAK-KEKAYAKAEAKAKAEAKAYKAAEAKKKAKAEAKKYAKA-AKAEKKEYAAAAE 58  
Db 2 AK 61

Qy 59 AKYKAEA-AKAAAEAA 74  
Db 62 AKAKAKAKAKAKAKA 78

RESULT 2  
US-08-167-641C-64  
; Sequence 64, Application US/08167641C  
; Patent No. 6033884  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/167,641C  
FILING DATE: December 14, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/855,389  
FILING DATE: March 20, 1992  
APPLICATION NUMBER: PCT/US93/02725  
FILING DATE: March 19, 1993

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 205/012  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: "Lys Ala" in positions 3 to 100 may be  
OTHER INFORMATION: present or absent.

US-08-167-641C-64

Query Match 39.5%; Score 144.5; DB 3; Length 100;  
Best Local Similarity 58.4%; Pred. No. 2.7e-07;  
Matches 45; Conservative 6; Mismatches 23; Indels 3; Gaps 3;

Qy 1 AKKYAK-KEKAYAKAEAKAKAEAKAYKAAEAKKKAKAEAKKYAKA-AKAEKKEYAAAAE 58  
Db 2 AK 61  
Qy 59 AKYKAEA-AKAAAEAA 74  
Db 62 AKAKAKAKAKAKAKA 78

RESULT 3  
US-08-460-971A-64  
; Sequence 64, Application US/08460971A  
; Patent No. 6150168  
; GENERAL INFORMATION:  
; APPLICANT: Woo, Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,971A  
FILING DATE: June 5, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/167,641  
FILING DATE: December 14, 1993  
APPLICATION NUMBER: 07/855,389  
FILING DATE: March 20, 1992  
APPLICATION NUMBER: PCT/US93/02725  
FILING DATE: March 19, 1993

ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 212/063  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: "Lys Ala" in positions 3 to 100 may be  
OTHER INFORMATION: present or absent.

US-08-460-971A-64

Query Match 39.5%; Score 144.5; DB 4; Length 100;



```
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-8
;
; Query Match 36.6%; Score 134; DB 4; Length 643;
; Best Local Similarity 52.0%; Pred. No. 1.8e-05;
; Matches 39; Conservative 9; Mismatches 23; Indels 4; Gaps 2;
;
; QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
; DB 490 ATKVAEAEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 546
;
; QY 60 KYKAAEAKAAKAAEA 74
; DB 547 TKVAEAEKQKAAEA 561
;
; RESULT 7
; US-08-216-894-10
; Sequence 10, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-10
;
; Query Match 35.5%; Score 130; DB 2; Length 472;
; Best Local Similarity 51.4%; Pred. No. 3.1e-05;
; Matches 38; Conservative 9; Mismatches 23; Indels 4; Gaps 2;
;
; QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
; DB 360 ATKVAEAEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 416
;
; QY 60 KYKAAEAKAAKAAEA 73
; DB 417 TKVAEAEKQKAAEA 430
;
; RESULT 8
; US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-10
;
; Query Match 35.5%; Score 130; DB 4; Length 472;
; Best Local Similarity 51.4%; Pred. No. 3.1e-05;
; Matches 38; Conservative 9; Mismatches 23; Indels 4; Gaps 2;
;
; QY 1 AKKYAKKEKAYAKKAAKAAKAYKAAEAKKKAKAAKAYAKAAK-AEKKEYAAAAA 59
```





```
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-041-889-27

Query Match 35.4%; Score 129.5; DB 3; Length 214;
Best Local Similarity 45.7%; Pred. No. 1.5e-05;
Matches 43; Conservative 4; Mismatches 28; Indels 19; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAEAK-----AYKAAEAKKKAKAEAKKYAKAAKAEKKEY 54
Db 111 AKKVAK--KAPAKKATKAATKAAATKAPAKKATKAPAKKATKAPAKKAVATKSPAKV 168
Qy 55 AAA-----EAKYKAEAAKAAKAEAYEA 77
Db 169 TKAVKTVAKSVKRAATKAPAKKAAKRPATKA 202

RESULT 12
US-08-849-2
; Sequence 2, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuguang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,849
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28 DECEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-889-27

Query Match 35.4%; Score 129.5; DB 3; Length 214;
Best Local Similarity 45.7%; Pred. No. 1.5e-05;
Matches 43; Conservative 4; Mismatches 28; Indels 19; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAEAK-----AYKAAEAKKKAKAEAKKYAKAAKAEKKEY 54
Db 111 AKKVAK--KAPAKKATKAATKAAATKAPAKKATKAPAKKATKAPAKKAVATKSPAKV 168
Qy 55 AAA-----EAKYKAEAAKAAKAEAYEA 77
Db 169 TKAVKTVAKSVKRAATKAPAKKAAKRPATKA 202
```

```
; MOLECULE TYPE: protein
US-08-346-849-2

Query Match 33.6%; Score 123; DB 1; Length 433;
Best Local Similarity 46.2%; Pred. No. 0.00013;
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

Qy 2 KKYAKKEKAYAKAEAKKAEAKKAYKAAEAKKKAKAEAKKYAKA-AKAEK---KEYAAA 57
Db 290 KEEKKEKERRRWEREGAGARAEAEAKAKAEAEAKAEAKAEAKAKADKKKAEAKA 349
Qy 58 EAKYKAEAAKAAKAEAY 75
Db 350 AKKONKRAIRNSAKEADY 367

RESULT 13
US-08-293-284A-2
; Sequence 2, Application US/08293284A
; Patent No. 5955343
; GENERAL INFORMATION:
; APPLICANT: Holmes, Todd
; APPLICANT: Zhang, Shuguang
; APPLICANT: Rich, Alexander
; APPLICANT: Dipersio, C. Michael
; APPLICANT: Lockshin, Curtis
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/293,284A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,326
; FILING DATE: 28-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-6008A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 33.6%; Score 123; DB 2; Length 433;
Best Local Similarity 46.2%; Pred. No. 0.00013;
Matches 36; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

Qy 2 KKYAKKEKAYAKAEAKKAEAKKAYKAAEAKKKAKAEAKKYAKA-AKAEK---KEYAAA 57
Db 290 KEEKKEKERRRWEREGAGARAEAEAKAKAEAEAKAEAKAEAKAKADKKKAEAKA 349
Qy 58 EAKYKAEAAKAAKAEAY 75
```



**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	164	40.1	421	2	JV0057	tola protein - Esc	
2	157.5	38.5	376	2	AG0592	tola protein [impo	
3	156.5	38.3	1701	2	T09127	probable erythrocy	
4	155	37.9	394	2	F90725	membrane spanning	
5	155	37.9	394	2	G85576	membrane spanning	
6	154	37.7	347	2	E83525	Tola protein PA097	
7	151	36.9	210	2	A25550	histone H1 - sea u	
8	149	36.4	231	2	S59589	histone H1 - Chlam	
9	147.5	36.1	388	2	AC0138	Tola colicin impor	
10	147	35.9	214	2	G70673	probable hupB - My	
11	144	35.2	211	2	A28100	histone H1-beta, e	
12	142.5	34.8	344	2	S34153	met101-1 protein -	
13	142	34.7	372	2	G64064	outer membrane int	
14	140.5	34.4	328	2	A44993	cytosolic repetiti	
15	140.5	34.4	1390	2	S51364	sperm tail-specifi	
16	138	33.7	217	2	A26721	histone H1-gamma,	
17	137.5	33.6	1128	2	T30296	R27-2 protein - Tr	
18	135	33.0	219	2	E60110	repetitive protein	
19	134	32.8	206	2	S09388	histone H1 - sea u	
20	133.5	32.6	384	2	B43592	outer membrane pro	
21	133.5	32.6	433	2	S25194	zuotin - yeast (sa	
22	132.5	32.4	206	1	HSTR1P	histone H1 - rainb	
23	132	32.3	248	1	HSUR1P	histone H1, gonada	
24	131	32.0	208	2	T23778	histone H1.1 - Cae	
25	130.5	31.9	220	2	A28456	histone H1.10 - ch	
26	130.5	31.9	243	2	AE1689	hypothetical prote	
27	129.5	31.7	924	2	T06636	hypothetical prote	
28	129.5	31.7	1403	2	T11583	probable translati	
29	129	31.5	218	2	A23055	histone H1.01 - ch	



A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-347 <STO>  
A;Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: toIA; PA0971

Query Match 37.7%; Score 154; DB 2; Length 347;  
Best Local Similarity 47.6%; Pred. No. 0.00021;  
Matches 49; Conservative 11; Mismatches 21; Indels 22; Gaps 4;

Qy 1 AKYAKKEKAYAKKAEAKKAEAKYAA-----EAKKAKAEAKYA----- 44  
Db 120 ARKAAQAAAEAKKAADEAKKAAAEQKQADIAKRAEDAEAKKAAEDAKKAAEDA 179  
Qy 45 --KAAKAEKCEYAAAEAKYA--EAKKAYKAEAAKAAKAA 83  
Db 180 KKAABEAKKAAAEAKKAAAEAKK--KAAAAAAAARAKAA 220

RESULT 7  
A25550  
histone H1 - sea urchin (*Lytechinus pictus*)  
C;Species: *Lytechinus pictus* (painted urchin)  
C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999  
C;Accession: A25550  
R;Knowles, J.A.; Childs, G.J.  
Nucleic Acids Res. 14, 8121-8133, 1986  
A;Title: Comparison of the late H1 histone genes of the sea urchins *Lytechinus pictus* and  
A;Reference number: A25550; MUID:87040778; PMID:3022245  
A;Accession: A25550  
A:Molecule type: DNA  
A;Residues: 1-210 <KNO>  
A;Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617  
A;Superfamily: histone H1  
C;Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 36.9%; Score 151; DB 2; Length 210;  
Best Local Similarity 51.7%; Pred. No. 0.00024;  
Matches 46; Conservative 10; Mismatches 25; Indels 8; Gaps 3;

Qy 2 KKYAKKEKAYAKKAEAKKAEAKYKAAEAK-KKAAEAKKYAKAAEAKKEYAAEAK 60  
Db 96 KTEAQKAAAKKAAKAAKKEQKQAKTKARKEKLAAKKAAKKA-AKKYKPAKAK 154  
Qy 61 YKAE-----AAKKAYKAEAAKAAKAA 83  
Db 155 KPAAKAAKPAKAAKAAKPAKPAKPAKAA 183

RESULT 8  
S59589  
histone H1 - *Chlamydomonas reinhardtii*  
C;Species: *Chlamydomonas reinhardtii*  
C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 23-Jul-1999  
C;Accession: S59589; S62122  
R;Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.  
Curr. Genet. 28, 333-345, 1995  
A;Title: The organization structure and regulatory elements of *Chlamydomonas* histone gen  
A;Reference number: S59581; MUID:96120862; PMID:8590479  
A;Accession: S59589  
A;Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A;Residues: 1-231 <PAB>  
A;Cross-references: EMBL:U16726  
A;Note: the authors did not translate the codon for residue 1  
R;Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
submitted to the EMBL Data Library, October 1994  
A;Description: The organization, structure and controlling elements of *Chlamydomonas* his  
A;Reference number: S62122  
A;Accession: S62122  
A:Molecule type: DNA

A;Residues: 1-173, 'P', 174-231 <FAW>  
A;Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98452.1; PID:g571480  
C;Genetics:  
A;Introns: 62/3; 101/3  
C;Superfamily: histone H1  
C;Keywords: chromosomal protein; DNA binding; nucleosome

Query Match 36.4%; Score 149; DB 2; Length 231;  
Best Local Similarity 39.5%; Pred. No. 0.00035;  
Matches 47; Conservative 7; Mismatches 29; Indels 36; Gaps 2;

Qy 1 AKYAKKEKAYAKKAEAKKAAE-----KAYKAAEAKKAKAEAKKYAKAAEAK 51  
Db 103 AKKAATPKAAAPKKEGAVKTKAPKKEGKKPKSAKAEKKPKKEGKKKAAKPAKAE 162  
Qy 52 KEYAA-----AEAKYKAEAAKAYKAEAAKAAKAA 83  
Db 163 KPAAKPAKTKTKAAAKPKAEKPKAAAKPKAEKPKAAKPAKAEKKAAPKAAKSA 221

RESULT 9  
AC0138  
TolA colicin import membrane protein [imported] - *Yersinia pestis* (strain CO92)  
C;Species: *Yersinia pestis*  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C;Accession: AC0138  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, I.  
Nature 413, 523-527, 2001  
A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AC0138  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-388 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC89966.1; PID:g15979190; GSPDB:GN00175  
C;Genetics:  
A;Gene: toIA

Query Match 36.1%; Score 147.5; DB 2; Length 388;  
Best Local Similarity 45.2%; Pred. No. 0.00067;  
Matches 47; Conservative 9; Mismatches 21; Indels 27; Gaps 3;

Qy 7 KEKAYAKKAEAKKAEAKYK-----AAEAKKKAKAEAKKYAKAAEAKKE----- 53  
Db 155 KAKEEQKAEATAAAQAKAEADKIVKQAEAKKAEAKKEAAVAAAKKQADADAKAV 214  
Qy 54 -----YAAEAKYKAEAAK-AYKAEAAKAAKAA 83  
Db 215 EVAEKAADAAEAKKAAADAEKKAATAAKKVAATAAEAKKAAAEAA 258

RESULT 10  
G70673  
probable hupB - *Mycobacterium tuberculosis* (strain H37RV)  
C;Species: *Mycobacterium tuberculosis*  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: G70673  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: G70673  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A;Residues: 1-214 <COL>  
A;Cross-references: GB:Z83018; GB:AL123456; NID:g3261671; PIDN:CAB05427.1; PID:g1694845  
A;Experimental source: strain H37RV  
C;Genetics:







THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 19.5949 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-6

Perfect score: 409

Sequence: 1 AKKYAKKEKAYAKKAERAAK.....KKAYKAEAAKAAKEAAAYEA 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	40.1	421	1 TOLA_ECOLI	P19934 escherichia
2	154	37.7	347	1 TOLA_PSEAE	P50600 pseudomonas
3	151	36.9	210	1 HI_LYTP1	P06144 lytechinus
4	147	35.9	214	1 DBH_MYCTU	P95109 mycobacteri
5	146.5	35.8	205	1 DBH_MYCBO	Q9xb18 mycobacteri
6	142.5	34.8	344	1 MST1_DROHY	Q08695 drosophila
7	142	34.7	372	1 TOLA_HAEIN	P44678 haemophilus
8	141	34.5	211	1 H1B_STRPU	P15869 strongyloce
9	140.5	34.4	1391	1 MST2_DROHY	Q08696 drosophila
10	138	33.7	217	1 H1G_STRPU	P07796 strongyloce
11	137.5	33.6	346	1 R56_AEDAE	Q9u761 aedes aegyp
12	137	33.5	349	1 R56_AEDAL	Q9u762 aedes albop
13	136.5	33.4	208	1 DBH_MYCSM	Q2zhc5 mycobacteri
14	133.5	32.6	384	1 TMPB_TREPH	P29720 treponema p
15	133.5	32.6	433	1 ZUO1_YEAST	P32527 saccharomyc
16	132.5	32.4	206	1 HI_ONCMY	P06350 oncorhynch
17	132	32.3	248	1 HI_PARAN	P02256 parorchinus
18	131	32.0	207	1 H11_CAEL	P10771 caenorhabdi
19	130.5	31.9	219	1 H110_CHICK	P08286 gallus gall
20	129.5	31.7	1403	1 XDF3_SCHPO	Q10475 schizosacch
21	129	31.5	241	1 H101_CHICK	P08284 gallus gall
22	127	31.1	139	1 ASR_KLEPN	Q93mh5 klebsiella
23	126.5	30.9	217	1 H1_AANPL	P09426 anas platyr
24	126.5	30.9	240	1 H12_VOLCA	Q08865 volvox cart
25	125.5	30.7	217	1 H11_CHICK	P09987 gallus gall
26	125.5	30.7	223	1 H103_CHICK	P08285 gallus gall
27	123.5	30.2	224	1 H11L_CHICK	P08287 gallus gall
28	122.5	30.0	219	1 H1B_XENLA	P06893 xenopus lae
29	121.5	29.7	237	1 H1E_CHITE	P40278 chironomus
30	121.5	29.7	238	1 H1_WHEAT	P27806 triticum ae
31	121	29.6	225	1 H15_HUMAN	P16401 homo sapien
32	120.5	29.5	771	1 CALD_CHICK	P12957 gallus gall
33	120	29.3	232	1 H1B_CHITE	P40276 chironomus

34	120	29.3	325	1	TMPE_TREPA	P19649 treponema p
35	118.5	29.0	194	1	H1_SALTR	P02254 salmo trutt
36	118	28.9	218	1	H1LR_CHICK	P08288 gallus gall
37	118	28.9	221	1	H1C_CHITE	P40277 chironomus
38	118	28.9	244	1	H1O_CHITH	Q07134 chironomus
39	117	28.6	218	1	H14_HUMAN	P10412 homo sapien
40	117	28.6	233	1	H1L_GLYSA	P40286 glyptotendi
41	116.5	28.5	833	1	IF2_PASMU	P57873 pasteurella
42	115.5	28.2	209	1	H1A_XENLA	P06892 xenopus lae
43	115.5	28.2	216	1	H1C1_XENLA	P15866 xenopus lae
44	115.5	28.2	233	1	H11_GLYBA	P40263 glyptotendi
45	115.5	28.2	239	1	RL23_DROME	P50887 drosophila

## ALIGNMENTS

RESULT 1  
TOLA\_ECOLI  
ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OC Escherichia  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JM105;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levengood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
RT their products, components of a multistep translocation system in  
RT Escherichia coli.";  
RL J. Bacteriol. 171:6600-6609(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MGL655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayaishi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levengood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "Tola: a membrane protein involved in colicin uptake contains an  
RT extended helical region.";  
RN Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
RN [5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

```

RX MEDLINE=97113525; PubMed=8955385;
RA Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT "Identification and characterization of the tolQRA genes of
RL Pseudomonas aeruginosa.";
RN J. Bacteriol. 178:7059-7068 (1996).
RP [2]
RA REVISIONS TO N-TERMINUS.
RA Duan K., Sokol P.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RL opportunistic pathogen."
RN Nature 406:959-964 (2000).
CC -I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39558; AAC44660.2; -.
DR EMBL; AB004530; AAG04360.1; -.
DR TR Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).
FT DOMAIN 209 216 POLY-ALA.
SQ SEQUENCE 347 AA; 37935 MW; EEDD4B04AA095945 CRC64;

Query Match 37.7%; Score 154; DB 1; Length 347;
Best Local Similarity 47.6%; Pred. No. 0.00011;
Matches 49; Conservative 11; Mismatches 21; Indels 22; Gaps 4;

QY 1 AKKYAKKEKAYAKKAEEKAEKAYKAA-----EAKKKAEAKKAYA--- 44
Db . : : | | | | | | | | | | | | | | | | | | | | | |
QY 45 --KAAKAEKEYAAAAEKYKA--EAAKYAKAEAKKAAKAAEA 83
Db 120 ARKAEQAQKAAEAKKADENKAEKAEKAAEQKKQNDIAKKRADEAKKAAEDAKKAAEDA 179
Db 180 KKKAEEAKKAAEAEEKKAAVEAAKK--KAAAAAAAAARKAA 230

RESULT 3
H1_LYTPI ID H1_LYTPI STANDARD; PRT; 210 AA.
AC P06144;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Late histone H1.
OS Lytechinus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euarchinoidea; Echinacea; Temnopneuroidea; Toxopneustidae;
OC Lytechinus.
OX NCBI_TaxID:7653;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Embryo;
RX MEDLINE=87040778; PubMed=302245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RL Lytechinus pictus and Strongylocentrotus purpuratus.";
EL Nucleic Acids Res. 14:8121-8133(1986).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04488; CAA28177.1; -.
DR PIR; A25550; A25550.
DR HSP; P02259; LHST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist N.
DR Pfam; PF00538; linker_histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;
-----
Query Match 36.9%; Score 151; DB 1; Length 210;
Best Local Similarity 51.7%; Pred. No. 0.00012;
Matches 46; Conservative 10; Mismatches 25; Indels 8; Gaps 3;
Oy 2 KYAKKEKAYAKKAAKAAKAYKAAKAAK-KKAAKAAKAYKAAKAAKAAKAAKAAK 60
Db 96 KTEAQKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 154
Oy 61 YKAE-----AAKAYKAAKAAKAAKAA 83
Db 155 KPAKAAKPAKAAKPAKAAKPAKAAK 103
-----
RESULT 4
DBH_MYCTU
ID DBH MYCTU STANDARD; PRT; 214 AA.
AC P95I09;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa
DE laminin-2-binding protein)
GN HUP OR HLP OR LB21 OR RV2986C OR MT3064 OR MTCY349.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaija F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares R., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 71-86, AND DNA-BINDING.
RX STRAIN=H37RV;
RA Prasad H.K., Annapurna P.S., Dey A.B., Tyagi J.S., Jain N.K.,
RA Savita P.;
RL Submitted (DEC-1997) to the SWISS-PROT data bank.
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z83018; CAB05427.1; -.
DR EMBL; AE007127; AAK47393.1; -.
DR HSP; P02346; LHUU.
DR TIGR; MT3064; -.
DR Tuberculist; RV2986C; -.
DR InterPro; IPR000119; Bac_DNABind.
DR InterPro; IPR001386; Histone H1/H5.
DR Pfam; PF00216; Bac DNA binding; 1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000945; Bac DNABind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
DR KW DNA-binding; DNA condensation; Repeat; Complete proteome.
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 101 214 DEGENERATE REPEATS REGION.
SQ SEQUENCE 214 AA; 21187 MW; CB09AF20FB353544 CRC64;
-----
Query Match 35.9%; Score 147; DB 1; Length 214;
Best Local Similarity 50.0%; Pred. No. 0.00023;
Matches 47; Conservative 3; Mismatches 34; Indels 10; Gaps 3;
Oy 1 AKKYAKKEKAYKAAKAAKAAKAAK-----AYKAAEAKKKAAKAAKAAKAAKAAK 54
Db 111 AKKVAK--KAPAKATKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 168
Oy 55 AAAEAK--YKAEAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 86
Db 169 TKAVKUTAVKAVKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 202
-----
RESULT 5
DBH_MYCBO
ID DBH MYCBO STANDARD; PRT; 205 AA.
AC Q9X518; Q9S5J5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN HUP OR HLP OR MDP1.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
```





KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.  
SQ SEQUENCE 217 AA; 22658 MW; C725IEED3413B185 CRC64;

Query Match 33.7%; Score 138; DB 1; Length 217;  
Best Local Similarity 46.2%; Pred. No. 0.0009;  
Matches 43; Conservative

QY 2 KKYAKKEKAYAKAEK-----AAKKAFAKAYKAAEAKKKAKAEAKKYAKA 49  
DB 125 KTKARKEKVAAKAAKATKTCTKVKVPAKAKKPAKKPAKKPAKKPAKKPAKKVAK 184  
QY 50 EKKEYAAAEKYYKAAEAAYKAAEAAKAAKEA 82  
DB 185 PAKKAAAPPAKKAAPKAKA--AKPAKKAAPKA 215

RESULT 11  
RS6\_AEDAE STANDARD; PRT; 346 AA.

ID RS6\_AEDAE AC Q9U761;  
DT 15-JUN-2002 (Rel. 41, Created)  
DY 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DD 40S ribosomal protein S6.  
GN RPS6.  
OS Aedes aegypti (Yellowfever mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
OC Culicoidae; Aedes.  
OX NCBI\_TaxID=7159;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT Gavino V.H., Fallon A.M.;  
RA "Aedes mosquitoes ribosomal protein S6 cDNA";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH  
AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR  
CLASSES OF MRNA (By similarity).  
CC -I- PTM: Ribosomal protein S6 is the major substrate of protein  
kinases in eukaryote ribosomes (By similarity).  
CC -I- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.ebi.ac.uk/ebis-sib.ch/  
or send an email to license@ebi-sib.ch).

EMBL: AF154067; AAF04790.1;  
DR InterPro: IPR001377; Ribosomal\_S6E.  
DR Pfam: PF01092; Ribosomal\_S6e; 1.  
DR ProDom: PD003460; Ribosomal\_S6e; 1.  
DR PROSITE: PS00578; RIBOSOMAL\_S6e; 1.  
KW Ribosomal protein; Phosphorylation.  
SQ SEQUENCE 346 AA; 39365 MW; 599CFF7B22BD8DFC CRC64;

Query Match 33.8%; Score 137.5; DB 1; Length 346;  
Best Local Similarity 51.1%; Pred. No. 0.0016;  
Matches 45; Conservative

QY 1 AKKYAKKE-KAYAKAEAKKAEAKYKAAEAKK-KAKAEAKKYAKAAEK-EYAAA 57  
DB 255 AKKVAKKEAKEKVVTEAAKKADAKAKTAEPKAEKSKSETGKTAGDKKEKVEKKA 314  
QY 58 EAKYKAEAAKKAYKAAEAAKAAEAAEAYE 85  
DB 315 PAVAKKEAPKR--KPEAAKGDAAGAAKE 340

RESULT 12  
RS6\_AEDAL



```

ID AC RSC_ABDAL STANDARD; PRT; 349 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S6.
OS R96.
GN Aedes albopictus (Forest day mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Aedes.
OX NCBI_TaxID=7160;
RN [1]
RP SEQUENCE FROM N.A.
RA Gavino V.H., Fallon A.M.;
RT "Aedes mosquitoes ribosomal protein S6 cDNA."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH
CC AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR
CC CLASSES OF MRNA (By similarity).
CC -!- PFM: Ribosomal protein S6 is the major substrate of protein
CC kinases in eukaryote ribosomes (By similarity).
CC -!- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; AF154066; AAF04789.1; -
CC DR InterPro; IPR001377; Ribosomal_S6E.
CC DR Pfam; PF01092; Ribosomal_S6E; 1.
CC DR ProDom; PD003460; Ribosomal_S6E; 1.
CC DR PROSITE; PS00578; RIBOSOMAL_S6E; 1.
CC DR Ribosomal protein; Phosphorylation.
CC KW RIBOSOMAL PROTEIN;
CC SEQUENCE 349 AA; 39619 MW; 669A5DA33F8BADD3 CRC64;
CC
CC Query Match 33.5%; Score 137; DB 1; Length 349;
CC Best Local Similarity 49.5%; Pred. No. 0.0017;
CC Matches 45; Conservative 5; Mismatches 33; Indels 8; Gaps 4;
CC
CC QY 1 AKYAKKE-KAYAKKAAKAAKAAKAYKAKA---AEAKKKAAKAAKAYKAAKAAKAYK- 54
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 255 AKYAKKAAKKEVKVYTAAKKADAKAAKAVKPKADKKSADSGKATAGDKKKVEK 314
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC QY 55 AAEEKYKAAKAYKAAKAAKAAKAAKAAE 85
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 315 KAAPAAKKEAPKR--KPEAAKGDSAAKKE 343
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC
CC RESULT 13
CC DBH_MYCSM
CC ID DBH_MYCSM STANDARD; PRT; 208 AA.
CC AC Q9ZHC5;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).
CC GN HUP OR HLP.
CC OS Mycobacterium smegmatis.
CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC OX NCBI_TaxID=1772;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA Lee B.H., Murugan-Oei B., Dick T.;
CC RX MEDLINE=99110209; PubMed=9894918;
CC RA Lee B.H., Murugan-Oei B., Dick T.;
CC RT "Upregulation of a histone-like protein in dormant Mycobacterium
CC smegmatis."

```

```

RL Mol. Gen. Genet. 260:475-479 (1998).
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; AF068138; AAD13809.1; -
CC DR HSP; P02346; IHU.
CC DR InterPro; IPR000119; Bac_DNABind.
CC DR Pfam; PF00216; Bac_DNA_Binding; 1.
CC DR ProDom; PD000945; Bac_DNABind; 1.
CC DR SMART; SM00411; BHL; 1.
CC DR PROSITE; PS00045; HISTONE LIKE; 1.
CC KW DNA-binding; DNA condensation; Repeat.
CC FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
CC FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
CC SQ SEQUENCE 208 AA; 21230 MW; CA5F577F61F7EF09 CRC64;
CC
CC Query Match 33.4%; Score 136.5; DB 1; Length 208;
CC Best Local Similarity 51.6%; Pred. No. 0.0012;
CC Matches 47; Conservative 6; Mismatches 29; Indels 9; Gaps 4;
CC
CC QY 1 AKYAKK---EKAYAKK-AEKAAKAAKAAK---AYKAAEKKKAAKAAKAYKAAKAAK- 51
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 111 AKKAAKAPAKKAAKAAKKTATKAAAKKAPAKKAAKAAKAAKAAKAAKAAKAA 170
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC QY 52 KEYAAAEAKYKAAKAAKAYKAAKAAKAAKAA 82
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 171 TKAPAKKAAKAAKAPAKKAAKAAKAAKAAKAA 201
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC
CC RESULT 14
CC TMPB_TREPH
CC ID TMPB_TREPH STANDARD; PRT; 384 AA.
CC AC P29720;
CC DT 01-APR-1993 (Rel. 25, Created)
CC DT 01-APR-1993 (Rel. 25, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Treponema membrane protein B precursor (Antigen tmpB).
CC GN TMPB.
CC OS Treponema phagedenis.
CC OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
CC OX NCBI_TaxID=162;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Kazan 5;
CC RX MEDLINE=91372983; PubMed=1894368;
CC RA Yelton D.B., Limberger R.J., Curci K., Malinosky-Rummell F.,
CC RA Slivinsky L., Schouls L.M., van Embden J.D., Charon N.W.;
CC RT "Treponema phagedenis encodes and expresses homologs of the Treponema
CC pallidum tmpA and tmpB proteins."
CC RL Infect. Immun. 59:3685-3693 (1991).
CC -!- FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR
CC LARGE MOLECULES.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: TO TMPB OF T.PALLIDUM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 45.5401 Seconds  
(without alignments)  
389.109 Million cell updates/sec

Title: US-09-816-989a-6  
Perfect score: 409  
Sequence: 1 AKKYAKKEKAYAKAEKAAK.....KAYKAEKAAKAEKAYEA 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 21:\*
- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	165.5	40.5	407	16	Q8ZQT6 salmonella
2	163	39.9	372	2	Q9WX1 pseudomonas
3	157.5	38.5	376	16	Q8Z8C1 salmonella
4	156.5	38.3	1701	5	O61164 plasmodium
5	155	37.9	394	16	Q8X965 escherichia
6	154	37.7	395	2	Q937K4 erwinia chr
7	151	36.9	1341	16	Q98KG7 rhirubium 1
8	148.5	36.3	232	10	Q39576 chlamydomon
9	148	36.2	389	16	Q9CM70 pasteurella
10	147.5	36.1	388	16	Q8ZGZ2 yersinia pe
11	146	35.7	1866	5	Q8T5C8 plasmodium
12	139	34.0	200	16	Q8XVN7 ralestonia s
13	137.5	33.6	1128	5	Q26947 trypanosoma
14	134.5	32.9	212	3	Q93946 candida alb
15	131.5	32.2	224	13	Q90ZD7 bufo bufo g
16	131	32.0	208	5	Q93901 caenorhabdi

17	130.5	31.9	243	16	Q92A67
18	130.5	31.9	275	5	O01395
19	129.5	31.7	485	10	Q8RXD0
20	129.5	31.7	924	10	Q9SU08
21	129	31.5	755	10	Q9FP71
22	128	31.3	1671	5	Q8T5C9
23	127.5	31.2	98	5	Q8WQ44
24	127.5	31.2	111	5	Q8T9R3
25	127.5	31.2	275	10	Q9XHL9
26	127.5	31.2	311	12	Q84528
27	127.5	31.2	619	5	Q95S18
28	126.5	30.9	284	10	Q65794
29	126	30.8	265	10	Q39598
30	126	30.8	288	10	Q65795
31	126	30.8	845	5	Q9YIP8
32	125	30.6	581	16	Q9RU45
33	124.5	30.4	909	10	Q9SU09
34	124.5	30.4	969	5	Q9NDI9
35	124	30.3	182	2	Q45370
36	124	30.3	298	2	Q52088
37	124	30.3	607	5	Q9V6S7
38	123.5	30.2	191	5	O46142
39	123	30.1	356	16	Q9XR10
40	122.5	30.0	191	5	O46141
41	122.5	30.0	237	10	Q9SWU2
42	122.5	30.0	237	10	Q9ZR20
43	122.5	30.0	295	10	Q9AT22
44	122.5	30.0	295	10	Q9AT22
45	122.5	30.0	297	10	Q9SXQ8

ALIGNMENTS

RESULT 1

Q8ZQT6	ID	Q8ZQT6	PRELIMINARY;	PRT;	407 AA.
AC	Q8ZQT6;				
DT	01-MAR-2002	(TrEMBLrel. 20, Created)			
DT	01-MAR-2002	(TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)			
DE	Toi protein, membrane spanning protein.				
GN	TOLA OR STM0747.				
OS	Salmonella typhimurium.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_TaxID=602;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;				
RA	"Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";				
RL	Nature 413:852-856(2001).				
DR	EMBL; AE008730; AAL19691.1; -.				
DR	InterPro; IPR00104; Antifreeze_1.				
DR	PRINTS; PR00308; ANTIFREEZE1.				
KW	Complete proteome.				
SQ	SEQUENCE 407 AA; 41865 MW; 2534352116602D75 CRC64;				

Query Match 40.5%; Score 165.5; DB 16; Length 407;  
Best Local Similarity 51.9%; Pred. No. 2.5e-05;  
Matches 55; Conservative 14; Mismatches 16; Indels 21; Gaps 6;

Qy	1	AKKYAKKEKAYAKAEKAA-----KKAEEKYK-AAEAKKKAKAEKKA-----KAAKAEK 51
Db	123	AAKLAQQQQQAEBAKAAADAKKKAEAEAAVKAADAKKKAEAE 182

```

Qy 52 KEYAAAEAKYKAE-----AKYKAEAKAAKAA---KEAAAYEA 86
ID Q9WXX1 PRELIMINARY; PRT; 372 AA.
Db 183 AK-AAADAKKAEAAKAAAEAKKAEAAEAAKAAAEAKKKADEA 227

RESULT 2
Q9WXX1
ID Q9WXX1 PRELIMINARY; PRT; 372 AA.
AC Q9WXX1;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Tola protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MT-2;
RX MEDLINE=96198174; PubMed=8626299;
RA Rodriguez-Herva J.J., Ramos-Consalvez M.I., Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
RT lipoprotein (PAL) is involved in maintenance of the integrity of the
RT cell envelope.";
RL J. Bacteriol. 178:1699-1706(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RA Ramos-Gonzalez I.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RA Rodriguez-Herva J.J.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MT-2;
RX MEDLINE=96422022; PubMed=8824639;
RA Rodriguez-Herva J.J., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RL J. Bacteriol. 178:5836-5840(1996).
DR EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR001386; Histone_H1/H5.
DR PRINTS; PR00624; HISTONEH5.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 39.9%; Score 163; DB 2; Length 372;
Best Local Similarity 53.1%; Pred. No. 3 5e-05;
Matches 52; Conservative 12; Mismatches 22; Indels 12; Gaps 5;

Qy 1 AKKYAKKEKA--YAKAE-KAAKAEAKAYKAA--BAKKKAEAKKYA-----KAAKA 49
Db 148 AKKAAEQADIAKKAEDEBAKKAAAEAKKAAAEAKKAAEDAKKAAEBAKKAED 207

Qy 50 EKEYAAAEAKYK-AEAKKAYKAEAKAAKAEAAEYEA 86
Db 208 AKKAAAEADAKKAAAEBAKKKAAADAQKKAQAEARKA 245

RESULT 3
Q8Z8C1 PRELIMINARY; PRT; 376 AA.
ID Q8Z8C1;
AC Q8Z8C1;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Tola protein.
GN STK0793.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627268; CAD05209.1; -.
DR InterPro; IPR00104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01574; TUBBYPROTEIN.
KW Complete proteome.
SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;

Query Match 38.5%; Score 157.5; DB 16; Length 376;
Best Local Similarity 54.8%; Pred. No. 9.3e-05;
Matches 51; Conservative 10; Mismatches 17; Indels 15; Gaps 5;

Qy 1 AKKYAKKEKAYKAEAKKAEAKYK-AEAKKAKAEAKKYA-----KAAKAEKKEYA 55
Db 127 AQOQQQAEEA-AKAAADAKKAEAEAAKAAADAKKAEAEAAKAAADAKKAEAEAAK-A 184

Qy 56 AAEAKYKAEAE-----AKKAYKAEAAKAAAK 80
Db 185 AAEAKKAEAEAAKAAADAKKADAEAAKAAAE 217

RESULT 4
Q61164 PRELIMINARY; PRT; 1701 AA.
ID Q61164;
AC Q61164;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MABEL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria
RT parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
DR EMBL; AF031886; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;

Query Match 38.3%; Score 156.5; DB 5; Length 1701;
Best Local Similarity 53.5%; Pred. No. 0.00048;
Matches 53; Conservative 9; Mismatches 18; Indels 19; Gaps 6;

Qy 1 AKKYAKKEKAYAE---KKAEEAKKAEAKYKAAEAKKKKA-----EAKYKAAK---- 48
Db 1204 AAKKAEERKKAEAVKKAEEAKKAEAEAKKAEAEAKKAEERKKAEAEAKKAEAEAKKAE 1262

Qy 49 AEKKEYAA--AEAKYKAEAAKKA-----YKAEAAKAAAKE 81
Db 1263 RKKAEAAKAEAEKKKAEAAKAEAEKKKAEAAKAEAE 1301

```

RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;  
 RL "Characterization of the *Erwinia chrysanthemi* *toi-pal* genes.";  
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ297885; CAC82708.1; -.  
 SQ SEQUENCE 395 AA; 41601 MW; 3C0C1DC12E181013 CRC64;

Query Match 37.7%; Score 154; DB 2; Length 395;  
 Best Local Similarity 53.3%; Pred. No. 0.00018;  
 Matches 49; Conservative 13; Mismatches 18; Indels 12; Gaps 5;

QY 1 AKKYAK--K EYAKAEKAAKAEKAYKAAEAKKAKAEAKKYAKAAEAKKEKAYAA 57  
 DB 166 AKQAESEVKKQAAADAKKAAEEBKTKAAABAAABAKKAAEEAK--AKAA-ADAKQAAEE 222  
 QY 58 EAYK-AEAAKAYKABAAK-----AAAKEAA 83  
 DB 223 EAKAKAAEAAKAAADAADAKKAAEAAAAAKKAA 254

RESULT 7  
 Q98KG7 PRELIMINARY; PRT; 1341 AA.  
 ID Q98KG7  
 AC Q98KG7;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein mll1482.  
 GS Mll1482.  
 GN Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OC NCBI\_TaxID=381;  
 RN NCBI\_TaxID=381;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT *Mesorhizobium loti*.";  
 RL DNA Res. 7:331-338(2000).  
 RT EMBL; AP002997; BAB48847.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1341 AA; 138670 MW; 79C5D8149C969BC8 CRC64;

Query Match 36.9%; Score 151; DB 16; Length 1341;  
 Best Local Similarity 52.2%; Pred. No. 0.00098;  
 Matches 47; Conservative 14; Mismatches 21; Indels 8; Gaps 5;

QY 1 AKKYAKKEKAYKAEKAAK-KAPAKYKAAEAKKAKAEAKKYAKA-----AKAEKEY 54  
 DB 1177 AEELQQEEAEARLKAEADADAKAPAEQAQAEAKQAQAEAKAKADAEAKAEAKAK 1236  
 QY 55 AAABAKYKAAEAKKAYKAE-AKAAAEKAA 83  
 DB 1237 ADAEAKAKADAEAKA-KADADAKAAAEQQA 1265

RESULT 8  
 Q39576 PRELIMINARY; PRT; 232 AA.  
 ID Q39576  
 AC Q39576;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Histone H1.  
 GN Chl.  
 GN Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.









GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 44.8143 Seconds  
(without alignments)  
255.712 Million cell updates/sec

Title: US-09-816-989a-6

Perfect score: 409

Sequence: 1 AKYAKKEKAYAKAEKAAK.....KKAYKAEAKAAKAEAYEA 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409	100.0	86	21 AAY82576	Copolymer molecule
2	351.5	85.9	77	21 AAY82575	Copolymer molecule
3	312.5	76.4	109	21 AAY82577	Copolymer molecule
4	268	65.5	66	21 AAY82574	Copolymer molecule
5	233	57.0	56	21 AAY82573	Copolymer molecule
6	163	39.9	154	11 AAR06445	Recombinant copoly
7	157	38.4	106	11 AAR06446	Recombinant copoly
8	156	38.1	100	21 AAY98499	Peptide #10 used i
9	156	38.1	100	21 AAY59044	Amino acid polymer
10	156	38.1	100	22 AAU04289	Poly-Lys-Ala used

11	156	38.1	100	22 AAB45852	Nucleic acid trans
12	147	35.9	214	20 AAY34055	M. tuberculosis hi
13	147	35.9	214	21 AAY57353	M. tuberculosis hi
14	146.5	35.8	205	21 AAB20575	Mycobacterium bovi
15	137.5	33.6	472	16 AAB84569	Trypanosoma cruzi
16	137.5	33.6	564	16 AAR84585	Trypanosoma cruzi
17	137.5	33.6	643	16 AAR84568	Trypanosoma cruzi
18	135	33.0	140	13 AAR25206	Synthetic helical
19	133.5	32.6	433	18 AAW30256	Zuotin. Saccharom
20	133.5	32.6	433	20 AAV32954	Human zuotin prote
21	133.5	32.6	433	22 AAG70739	S cerevisiae apopt
22	129.5	31.7	223	20 AAY14928	Amino acid sequenc
23	129	31.5	219	21 AAY76981	Quail Hi histone p
24	127.5	31.2	1507	21 AAB24128	Plasmodium yoelii
25	127	31.1	222	18 AAW29476	Human histone H1 i
26	127	31.1	222	20 AAV34033	Human histone H1 i
27	127	31.1	222	21 AAV57331	Human histone H1 i
28	126.5	30.9	45	21 AAY82572	Copolymer molecula
29	124	30.3	334	22 AAG28693	Novel human diagno
30	124	30.3	607	22 ABB62173	Drosophila melanog
31	123	30.1	130	23 AAU09945	Peptide fragment o
32	123	30.1	130	23 AAE13237	Human linker histo
33	121	29.6	158	20 AAV34068	Histone H1 isoform
34	121	29.6	158	21 AAY57366	Human histone H1.5
35	121	29.6	226	20 AAY34060	Human histone H1 i
36	121	29.6	226	21 AAY57358	Human histone H1 i
37	121	29.6	239	23 ABB49123	Listeria monocytog
38	119.5	29.2	332	21 AAY75540	Neisseria meningit
39	119.5	29.2	332	21 AAY75541	Neisseria meningit
40	119	29.1	234	23 AAU09944	Histone H1-4, frag
41	119	29.1	234	23 AAE13234	Human linker histo
42	118.5	29.0	111	18 AAW44829	Adhesive polyptpti
43	117	28.6	218	18 AAW29477	Human histone H1 i
44	117	28.6	218	20 AAV34034	Human histone H1 i
45	117	28.6	218	21 AAY57332	Human histone H1 i

#### ALIGNMENTS

RESULT 1

AAY82576

ID AAY82576 standard; peptide; 86 AA.

XX AAY82576;

XX AC

XX DT 28-JUL-2000 (first entry)

XX DE

XX DX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;

XX KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;

XX KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;

XX KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;

XX KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;

XX KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;

XX KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;

XX KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

XX KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

XX KW pemphigus vulgaris; systemic lupus erythematosus.

XX Unidentified.

XX OS

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PA (TEVA-) TEVA PHARM USA INC.





Db 50 -----YKAEAAKAAKEAAEYEA 66

RESULT 5  
AA082573  
ID AA082573 standard; peptide; 56 AA.  
AC AA082573;  
XX  
XX 28-JUL-2000 (first entry)  
XX  
XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.  
XX  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KW glatiramer acetate; autoimmune disease; antiarthritis; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;  
KW antinaeemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KW pemphigus vulgaris; systemic lupus erythematosus.  
XX  
XX Unidentified.  
XX OS  
XX WO200018794-A1.  
XX PN  
XX 06-APR-2000.  
XX PD  
XX 24-SEP-1999; 99WO-US22402.  
XX PF  
XX 25-SEP-1998; 98US-0101693.  
XX PR  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX PA (TEVA-) TEVA PHARM USA INC.  
XX XX  
XX Gad A, Lis D;  
XX PI  
XX WPI; 2000-317499/27.  
XX DR  
XX  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
PT glatiramer acetate and for treatment and prevention of immune diseases  
PT  
XX  
XX Claim 10; Page 14; 72pp; English.  
XX  
XX AA082571 to AA082577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular  
CC weight of a copolymer (CP), which has an identified molecular weight  
CC and an amino acid composition corresponding to the copolymer. The  
CC polypeptides of the invention are used as molecular weight markers for  
CC glatiramer acetate related tetrapolymers. The polypeptides may also be  
CC used for treating and preventing immune diseases in a mammal. Autoimmune  
CC diseases which may be treated include either cell-mediated or  
CC antibody-mediated diseases. Such diseases include arthritic conditions,  
CC demyelinating diseases and inflammatory conditions, e.g. multiple  
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic  
CC anaemia, autoimmune ophoritis, autoimmune thyroiditis, autoimmune  
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia  
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves  
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic  
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic  
CC lupus erythematosus. Mediated-mediated diseases which can be treated  
CC include host-versus-graft disease, graft-versus-host disease, and  
CC delayed-type hypersensitivity. The polypeptides of the invention have  
CC defined molecular weights and physical properties which are analogous to  
CC glatiramer acetate molecules, which makes them ideal for use as  
XX molecular weight markers.  
XX  
XX Sequence 56 AA;

Query Match 57.0%; Score 233; DB 21; Length 56;

Best Local Similarity 62.8%; Pred. No. 1.6e-14;  
Matches 54; Conservative 2; Mismatches 0; Indels 30; Gaps 1;

QY 1 AKKYAKKEKAYAKAEKAAKAEKAYKAAEAKKKAKAEAKKAKAEKKEKAEAAAEAK 60  
|||  
DB 1 AKKYAKKEKAYAKAEKAAKAEKAYKAAEAKKKAKAEAKKAKAEKKEKAEAAAEAK 60  
|||  
QY 61 YKAEAAKAYKAEKAAKAEKAAEAAEYEA 86  
|||  
DB 40 -----YKAEAAKAAKEAAEYEA 56  
|||

RESULT 6  
AA06445  
ID AAR06445 standard; protein; 154 AA.  
XX  
AC AAR06445;  
XX  
DT 03-JAN-1991 (first entry)  
XX  
DE Recombinant copolymer 1-77, myelin basic protein analogue.  
XX  
KW Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
KW immunological activity; autoimmune encephalomyelitis;  
KW multiple sclerosis;  
XX  
OS Synthetic.  
XX  
XX EP383620-A.  
XX PN  
XX 22-AUG-1990.  
XX PD  
XX 16-FEB-1990; 90EP-0301700.  
XX PF  
XX 07-FEB-1990; 90US-0473845.  
XX PR  
XX 17-FEB-1989; 89US-0312541.  
XX PR  
XX (REPL-) REPLIGEN CORP.  
XX PA  
XX Cook KS;  
XX PI  
XX WPI; 1990-255948/34.  
XX DR  
XX N-PSDB; AAQ05664.  
XX DR  
XX Producing genes encoding random polymers of aminoacid(s) - for  
PT producing recombinant polypeptide(s) with biological and/or  
PT immunological activity  
PT  
XX  
XX Disclosure; Fig 11; 25pp; English.  
XX  
XX To improve the expression of rCOP-1 polypeptides in E. coli, genes  
CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pEG3-2deltaN  
CC (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to  
CC express Protein A. The resulting plasmids encode fusion proteins  
CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.  
CC A methionine residue occurs between the Protein A and rCOP-1  
CC sequences, originating from the 5' linker sequence in order that  
CC the COP-1 polypeptide may be cleaved from the fusion protein.  
CC rCOP-1-77 contains oligonucleotide duplexes encoding the following  
CC segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue  
CC is left behind following CNBr cleavage of the fusion protein.  
CC The product prevents or arrests experimental autoimmune  
CC encephalomyelitis. They are used to prevent, arrest or control a  
CC demyelinating disorder, e.g. multiple sclerosis. They may also  
CC be used as additives to hair care products to confer beneficial  
CC effects on damaged hair or as supplements for diets deficient in  
CC certain amino acids.  
CC See also AAQ05665.  
XX  
XX Sequence 154 AA;

Query Match 39.9%; Score 163; DB 11; Length 154;  
Best Local Similarity 42.3%; Pred. No. 1e-07;













THIS PAGE BLANK (USPTO)



```

RESULT 4
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARKER
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match          55.5%; Score 268; DB 10; Length 66;
Best Local Similarity 75.6%; Pred. No. 1.2e-16;
Matches 65; Conservative 0; Mismatches 1; Indels 20; Gaps 3;

QY 1 AKYAKKEKAYAKKAAKKAEEKAYKAAKAAKKKAAKAAKKAEEKYKAAKAAKKEKYAAAAEAK 60
Db 1 AKYAKKEKAYAK-----AKKAAK-----AAKKAEEKYKAAKAAKKEKYAAAAEAK 49

QY 61 YKAAEAKKAYKAAEAKAAKAAKEAYEA 86
Db 50 -----YKAAEAKAAKAAKEAYEA 66

RESULT 5
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARKER
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

Query Match          57.0%; Score 233; DB 10; Length 56;
Best Local Similarity 62.8%; Pred. No. 9e-14;
Matches 54; Conservative 2; Mismatches 0; Indels 30; Gaps 1;

QY 1 AKYAKKEKAYAKKAAKKAEEKAYKAAEAKKKKAAEAKKYKAAKAAKKEKYAAAAEAK 60

```







Search completed: March 10, 2003, 12:53:44  
Job time : 25.7251 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 15.2405 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-6

Perfect score: 409

Sequence: 1 AKYAKKEKAYAKAEKAAK.....KKYAKBAKAAKAAKEAYEA 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	156	38.1	100	2	US-08-460-890A-64
2	156	38.1	100	3	US-08-167-641C-64
3	156	38.1	100	4	US-08-460-971A-64
4	156	38.1	100	4	US-08-462-040-64
5	147	35.9	214	3	US-09-041-889-27
6	137.5	33.6	472	2	US-08-216-894-10
7	137.5	33.6	472	4	US-09-115-746-10
8	137.5	33.6	564	2	US-08-216-894-2
9	137.5	33.6	564	4	US-09-115-746-2
10	137.5	33.6	643	2	US-08-216-894-8
11	137.5	33.6	643	4	US-09-115-746-8
12	133.5	32.6	433	1	US-08-346-849-2
13	133.5	32.6	433	2	US-08-293-284A-2
14	129.5	31.7	223	4	US-09-095-855-201
15	129.5	31.7	223	4	US-09-205-426-201
16	127.5	31.2	1507	3	US-08-929-329-5
17	127	31.1	222	3	US-09-041-889-3
18	127	31.1	222	3	US-08-837-058-3
19	121	29.6	158	3	US-09-041-889-40
20	121	29.6	226	3	US-09-041-889-32
21	119.5	29.2	56	4	US-08-993-008A-6
22	117	28.6	218	3	US-09-041-889-4
23	117	28.6	218	3	US-08-837-058-4
24	115.5	28.2	262	1	US-08-403-379A-1
25	115.5	28.2	262	2	US-08-929-414-1
26	115.5	28.2	263	2	US-08-557-309B-51
27	115.5	28.2	263	3	US-08-834-306-51

28	115.5	28.2	263	4	US-08-993-674A-51
29	115.5	28.2	263	4	US-09-256-976-51
30	109	26.7	100	2	US-08-460-890A-62
31	109	26.7	100	3	US-08-167-641C-62
32	109	26.7	100	4	US-08-460-971A-62
33	109	26.7	100	4	US-08-462-040-62
34	109	26.7	434	1	US-08-097-830E-3
35	109	26.7	434	2	US-08-456-112B-3
36	106.5	26.0	48	4	US-08-993-008A-5
37	105.5	25.8	219	2	US-08-557-309B-54
38	105.5	25.8	442	3	US-08-834-306-52
39	105.5	25.8	442	4	US-08-993-674A-52
40	105.5	25.8	442	4	US-09-256-976-52
41	105	25.7	92	4	US-09-344-529-2
42	105	25.7	1507	6	5268270-2
43	104	25.4	288	3	US-08-312-949-4
44	104	25.4	288	3	US-08-446-201-4
45	104	25.4	289	1	US-08-072-070-4

## ALIGNMENTS

### RESULT 1

US-08-460-890A-64

; Sequence 64, Application US/08460890A

; Patent No. 5994109

; GENERAL INFORMATION:

; APPLICANT: Woo, Savio L.C.

; APPLICANT: Smith, Louis C.

; APPLICANT: Cristiano, Richard J.

; APPLICANT: Gottchalk, Stephen

; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND

; TITLE OF INVENTION: METHODS OF USE

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; STREET: Suite 4700

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSeq for Windows 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,890A

; FILING DATE: June 5, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/167,641

; FILING DATE: December 14, 1993

; APPLICATION NUMBER: 07/855,389

; FILING DATE: March 20, 1992

; APPLICATION NUMBER: PCT/US93/02725

; FILING DATE: March 19, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard J.

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 212/066

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 64:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 100 amino acids

; TYPE: amino acid

; STRANDEDNESS: single





```

1  APPLICANT: Otsu, Keiko
2  TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
3  TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
4  NUMBER OF SEQUENCES: 10
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Foley & Lardner
7  STREET: 3000 K Street, N.W., Suite 500
8  CITY: Washington, D.C.
9  COUNTRY: USA
10 ZIP: 20007-5109
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/216,894
18 FILING DATE: 24-MAR-1994
19 ATTORNEY/AGENT INFORMATION:
20 NAME: BENT, Stephen A.
21 REGISTRATION NUMBER: 29,768
22 REFERENCE/DOCKET NUMBER: 85326/102/DRLO
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (202)672-5300
25 TELEFAX: (202)672-5399
26 TELEX: 904136
27 INFORMATION FOR SEQ ID NO: 10:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 472 amino acids
30 TYPE: amino acid
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33
34 US-08-216-894-10

```

Query Match 33.6%; Score 137.5; DB 2; Length 472;  
Best Local Similarity 48.9%; Pred. No. 9.5e-06;  
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy	1	AKYAKGEKAVAKAEKAAKAAEKAYKAAAEKKAKEAKKYAKAAK-AEKKEYAAAAA	59
Db	346	ATKVAAEAK--OKAAEATKVAAEAKQKAAEATKVAAEAKQKAAEA	402
Qy	60	KYKAAEAAK-KAYKAAEAAKAAKAAEAA	86
Db	403	TKVAAEAKQKAAEATKVAAEAKQKAAEA	430

```

RESULT 7
US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR
; TITLE OF INVENTION: WITH TRYPANOSOMA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, V
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894

```

FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 472 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-115-746-10

Query Match 33.6%; Score 137.5; DB 4; Length 472;  
Best Local Similarity 48.9%; Pred. No. 9.5e-06;  
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

[illegible]

RESULT 8  
US-08-216-894-2  
; Sequence 2, Application US/08216894

PATENT NO. 5076754  
 GENERAL INFORMATION:  
 APPLICANT: Kirchhoff, Louis V.  
 APPLICANT: Otsu, Keiko  
 TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
 TITLE OF INVENTION: WITH TRYPAÑOSOMA CRUZI  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Hardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/216,894  
 FILING DATE: 24-MAR-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 564 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-216-894-2

Query Match 33.6%; Score 137.5; DB 2; Length 564;  
Best Local Similarity 48.9%; Pred. No. 1.1e-05;  
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

```
Qy 1 AKKYAKKEKAYAKAEAKKAAKAYKAAAEAKKAAKAAKAYKAAK-AEKKEVAAAEA 59
Db 434 ATKVAEAEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 490
Qy 60 KYKAEAAK-KAYKAEAAKAAKAAKAAEA 86
Db 491 TKVAEAEKQKAAEATKVAEAEKQKAAEA 518

RESULT 9
US-09-115-746-2
; Sequence 2, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE: 08/216,894
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-2
Query Match 33.6%; Score 137.5; DB 4; Length 564;
Best Local Similarity 48.9%; Pred. No. 1.1e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAAKAYKAAAEAKKAAKAAKAYKAAK-AEKKEVAAAEA 59
Db 434 ATKVAEAEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 490
Qy 60 KYKAEAAK-KAYKAEAAKAAKAAKAAEA 86
Db 491 TKVAEAEKQKAAEATKVAEAEKQKAAEA 518

RESULT 10
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; FILING DATE:
```

```
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8
Query Match 33.6%; Score 137.5; DB 2; Length 643;
Best Local Similarity 48.9%; Pred. No. 1.3e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy 1 AKKYAKKEKAYAKAEAKKAAKAYKAAAEAKKAAKAAKAYKAAK-AEKKEVAAAEA 59
Db 462 ATKVAEAEK--QKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 518
Qy 60 KYKAEAAK-KAYKAEAAKAAKAAKAAEA 86
Db 519 TKVAEAEKQKAAEATKVAEAEKQKAAEA 546

RESULT 11
US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
```

```
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BENT, Stephen A.
/ REGISTRATION NUMBER: 29,768
/ REFERENCE/DOCKET NUMBER: 85326/102/DRLO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202)672-5300
/ TELEFAX: (202)672-5399
/ TELEX: 904136
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 643 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-09-115-746-8

Query Match 33.6%; Score 137.5; DB 4; Length 643;
Best Local Similarity 48.9%; Pred. No. 1.3e-05;
Matches 43; Conservative 11; Mismatches 29; Indels 5; Gaps 3;

Qy 1 AKKYAKKEKAYAKKAAKAAKAYKAAAEAKKAKAAKAYKAAK-AEKKEYAAEA 59
Db 462 ATKVAEEK---QAAEATKVAEAKQKAAEATKVAEAKQKAAEATKVAEAKQKAAEA 518

Qy 60 KYKAAAK-KAYKAAAKAAKAAEA 86
Db 519 TKVAEAKQKAAEATKVAEAKQKAAEA 546

RESULT 12
US-08-346-849-2
/ Sequence 2, Application US/08346849
/ Patent No. 5670483
/ GENERAL INFORMATION:
/ APPLICANT: Zhang, Shuguang
/ APPLICANT: Lockshin, Curtis
/ APPLICANT: Rich, Alexander
/ APPLICANT: Holmes, Todd
/ TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
/ TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
/ NUMBER OF SEQUENCES: 64
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
/ STREET: Two Militia Drive
/ CITY: Lexington
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02173-4799
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/346,849
/ FILING DATE: 28-DEC-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/973,326
/ FILING DATE: 28-DEC-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brook, David E.
/ REGISTRATION NUMBER: 22,592
/ REFERENCE/DOCKET NUMBER: MIT-6008
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 861-6240
/ TELEFAX: (617) 861-9540
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 433 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 32.6%; Score 133.5; DB 2; Length 433;
Best Local Similarity 47.6%; Pred. No. 2e-05;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Qy 2 KKYAKKEKAYAKKAAKAAKAYKAAAEAKKAKAAKAYKAAK-AKAEKKEYAAEA 60
Db 290 KEEKKEKERRKWEREAGARAEAEAKAKAEAEAKAESEAKANASAKADKK-----KAK 344

Qy 61 YKAEAAKAYKAAKAAKAAEA 84
Db 345 EAAKAAKKNK-RAIRNSAKEADY 367

/ MOLECULE TYPE: protein
US-08-346-849-2

Query Match 32.6%; Score 133.5; DB 1; Length 433;
Best Local Similarity 47.6%; Pred. No. 2e-05;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Qy 2 KKYAKKEKAYAKKAAKAAKAYKAAAEAKKAKAAKAYKAAK-AKAEKKEYAAEA 60
Db 290 KEEKKEKERRKWEREAGARAEAEAKAKAEAEAKAESEAKANASAKADKK-----KAK 344

Qy 61 YKAEAAKAYKAAKAAKAAEA 84
Db 345 EAAKAAKKNK-RAIRNSAKEADY 367

/ MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 32.6%; Score 133.5; DB 2; Length 433;
Best Local Similarity 47.6%; Pred. No. 2e-05;
Matches 40; Conservative 12; Mismatches 25; Indels 7; Gaps 3;

Qy 2 KKYAKKEKAYAKKAAKAAKAYKAAAEAKKAKAAKAYKAAK-AKAEKKEYAAEA 60
Db 290 KEEKKEKERRKWEREAGARAEAEAKAKAEAEAKAESEAKANASAKADKK-----KAK 344

Qy 61 YKAEAAKAYKAAKAAKAAEA 84
```

Db 345 EAAKAAKKKKNK-RAIRNSAKEADY 367

## RESULT 14

US-09-855-201

; Sequence 201, Application US/09095855

; Patent No. 6160093

; GENERAL INFORMATION:

; APPLICANT: Tan, Paul

; APPLICANT: Visser, Elizabeth

; APPLICANT: Skinner, Margot

; APPLICANT: Prestidge, Ross

; TITLE OF INVENTION: Compounds and Methods for

; TITL OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections

; NUMBER OF SEQUENCES: 208

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Law Offices of Ann W. Speckman

; STREET: 2601 Elliott Avenue, Suite 4185

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98121

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/095,855

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/705,347

; FILING DATE: 29-AUG-1996

; APPLICATION NUMBER: 08/873,970

; FILING DATE: 12-JUN-1997

; APPLICATION NUMBER: 08/997,362

; FILING DATE: 23-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Sleath, Janet

; REGISTRATION NUMBER: 37,007

; REFERENCE/DOCKET NUMBER: 11000.1002c3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-269-0565

; TELEFAX: 206-269-0563

; TELEX:

; INFORMATION FOR SEQ ID NO: 201:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 223 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-855-201

Query Match 31.7%; Score 129.5; DB 4; Length 223;

Best Local Similarity 54.3%; Pred. No. 2.4e-05;

Matches 50; Conservative 5; Mismatches 26; Indels 11; Gaps 6;

Qy 1 AKKYAKK---EKAYAKKAEKAAKAEAK--AYKAAEAKKKAKAEAKKYAKAAKAEKKEYA 55

Db 112 ARKAAKAPAKKAAKAAKKA-APAKAPAKKAAATKAAPAKKATAAKKAAAPAKKATAAKK--- 167

Qy 56 AAEAKYKAEAAAKKAYK-AEAAKAAKAEAAEYEA 86

Db 168 AAPAK-KAPAKKAAATKAAPAKKAPAKKAPAKKATKA 198

## RESULT 15

US-09-426-201

; Sequence 201, Application US/09205426

; Patent No. 6406704

; GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Tan, Paul L. J.

; TITLE OF INVENTION: Diagnosis and Methods for Treatment and

; TITL OF INVENTION: Diagnosis of Mycobacterial Infections

; FILE REFERENCE: 11000.1002c4

; CURRENT APPLICATION NUMBER: US/09/205,426

; CURRENT FILING DATE: 1998-12-04

; EARLIER APPLICATION NUMBER: 09/095,855

; EARLIER FILING DATE: 1998-06-11

; EARLIER APPLICATION NUMBER: 08/997,362

; EARLIER FILING DATE: 1997-12-23

; EARLIER APPLICATION NUMBER: 08/873,970

; EARLIER FILING DATE: 1997-06-12

; EARLIER APPLICATION NUMBER: 08/705,347

; EARLIER FILING DATE: 1996-08-29

; NUMBER OF SEQ ID NOS: 208

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 201

; LENGTH: 223

; TYPE: PRT

; ORGANISM: Mycobacterium vaccae

; US-09-205-426-201

Query Match 31.7%; Score 129.5; DB 4; Length 223;

Best Local Similarity 54.3%; Pred. No. 2.4e-05;

Matches 50; Conservative 5; Mismatches 26; Indels 11; Gaps 6;

Qy 1 AKKYAKK---EKAYAKKAEKAAKAEAK--AYKAAEAKKKAKAEAKKYAKAAKAEKKEYA 55

Db 112 ARKAAKAPAKKAAKAAKKA-APAKAPAKKAAATKAAPAKKATAAKKATAAKKATAAKK--- 167

Qy 56 AAEAKYKAEAAAKKAYK-AEAAKAAKAEAAEYEA 86

Db 168 AAPAK-KAPAKKAAATKAAPAKKAPAKKAPAKKATKA 198

Search completed: March 10, 2003, 12:30:03

Job time : 15.2405 secs

THIS PAGE BLANK (USPTO)





```
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
A:Accession: F90725
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; UID:21156231; PMID:11258796
A:Accession: F90725
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034197.1; PID:gl3360233; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: EC0774

Query Match 35.0%; Score 181.5; DB 2; Length 394;
Best Local Similarity 52.2%; Pred. No. 4.7e-05;
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

QY 1 AKYAKKAE----KAYAKKAAKAEKKAYAKKEAKYAKAEAKKKAKAEAKKYAK-BAAK 55
DB 120 AEEAAQAEELKQKQAEAAKAAADAKAKAEADDKA--AEEAAKAAADAKKKAEBAAK 177
QY 56 AKKEAY-KAEAKKYAKAKAEKKEVAAAEAKK---AEAAYKAEAAKAAKEAA 106
DB 178 AAEEAQKAEAAALKKKAEAAEAAAEARKKAAAEKAAADKKAEAKAAEAKAA 232

RESULT 3
G85576
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
A:Accession: G85576
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; UID:21074935; PMID:11206551
A:Accession: G85576
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <STO>
A:Cross-references: GB:AE005174; NID:gl2513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:209
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: tolA

Query Match 35.0%; Score 181.5; DB 2; Length 394;
Best Local Similarity 52.2%; Pred. No. 4.7e-05;
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

QY 1 AKYAKKAE----KAYAKKAAKAEKKAYAKKEAKKAKAEAKKKAKAEAKKYAK-BAAK 55
DB 120 AEEAAQAEELKQKQAEAAKAAADAKAKAEADDKA--AEEAAKAAADAKKKAEBAAK 177
QY 56 AKKEAY-KAEAKKYAKAKAEKKEVAAAEAKK---AEEAAKAYKAEAAKAAKEAA 106
DB 178 AAEEAQKAEAAALKKKAEAAEAAAEARKKAAAEKAAADKKAEAKAAEAKAA 232

RESULT 4
A25550
histone H1 - sea urchin (Lytechinus pictus)
C:Species: Lytechinus pictus (painted urchin)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
A:Accession: A25550
R:Knowles, J.A.; Childs, G.J.
Nucleic Acids Res. 14, 8121-8133, 1986
A:Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus an
```

```
A:Reference number: A25550; UID:87040778; PMID:3022245
A:Accession: A25550
A:Molecule type: DNA
A:Residues: 1-210 <KNO>
A:Cross-references: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617
C:Superfamily: histone H1
C:Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 34.9%; Score 181; DB 2; Length 210;
Best Local Similarity 50.0%; Pred. No. 3.2e-05;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

QY 2 KKYAKAEKAYAKKAAKAEKKAYAKKEAKYAKAAE----AKKAKAEAKKYAKAEAKAK 57
DB 96 KTEAQKA-RAAAKKAALAAKKKEQKKAATKARKEKLAAKKAAKAAKVKVPPAAKAK 154
QY 58 KEAYKAEKAYAKAEKKEVAAAEAKK-----EAAKAYKAEAAKAAKEAA 106
DB 155 KPAKGAKKPPAAK--KAAKKPPAAKPPAKKAAKPPAAKKAAPKAAKPPAAKKA 208

RESULT 5
E83525
TolA protein PA0971 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
A:Accession: E83525
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; UID:20437337; PMID:10984043
A:Accession: E83525
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-347 <STO>
A:Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AAG04360.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: tolA; PA0971

Query Match 34.5%; Score 179; DB 2; Length 347;
Best Local Similarity 49.1%; Pred. No. 6.1e-05;
Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;

QY 6 KKAEEKYAKKAAKAEK-----KAYAKKAEKAYKAAEAKKKAKAEAKKYAKAEAKAKE 59
DB 99 OKLEQQQVAAAKAAEQKKADEARKAEAKAEAKKAEKAEKAEKAEKAEKQKQADIACKR 158
QY 60 AYKAEAKKYAKKAEKKEVAAAEAKKAEAKKAYKAEAAKAAKAAKAAKEAAVEA 109
DB 159 A-BDEAKK--KAAEDAKKK--AAEDAKKAAAEAKKAAAEAAKAAAEAAKAAVEA 204

RESULT 6
AG0592
tolA protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
A:Accession: AG0592
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
.; S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AG0592
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <PAR>
```



**QY** 5 AKKAEKAYAKKAKEKKAYAKKEAKAYKAAEAKKKKAAEAKKYAKEAFAAKKKEAYKA- 63  
Db 1205 AKKAE-E-RKKAEAVK-KAEAEAKKAAEAKKAAEKERKK-KAEAANKALERRKKKSAAKKAL 1261

**QY** 64 EAKYAKAKAKEKVEYAAEAACKAAEAKYKAEAKAARAKE 104  
Db 1262 ERKKKAAEAKKAEKKKAAEAKKAAEAEKK-KAEAAKKAEEE 1301

**RESULT 13**

A28456 histone H1.10 - chicken  
C/Species: Gallus gallus (chicken)  
C/Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 04-Mar-2000  
C/Accession: A28456  
R/Coles, L.S.; Robins, A.J.; Madley, L.K.; Wells, J.R.E.  
J Biol. Chem. 262, 9656-9663, 1987  
A/Title: Characterization of the chicken histone H1 gene complement. General  
A/Reference number: A92655; MUID:87250632; PMID:3597432  
A/Accession: A28456  
A/Molecule type: DNA  
A/Residues: 1-220 <COL>  
C/Cross-references: GB:M17018; NID:g211834; PIDN:AAA48788.1; PID:g211835  
C/Superfamily: histone H1  
C/Keywords: acetylated amino end; chromosomal protein; DNA binding; nucleosom  
F/2/Modified site: acetylated amino end (Ser) (in mature form) #status pred

**QY** 1 AKKYAKKAPKAYAKKAAKAEKKAYAKKEA----KAYKAAEAK---KKAAEAKKKYAKEA 53  
Db 109 SKKPGVGEKAPRKRTTPAAKPKKPAKKPASAANKPKKAAAKKSPPKAKKPPAAAAATKKA 168

**QY** 54 AKAKKEAYKAEAKKYAKAKAEKKEYAA-AEAKKAEAKAYKAEAAKAAAKE 104  
Db 169 AKSPKATTAAPKKMAATAKGPAAKAVKPKAAKPKAAKPKAAKAKKAAAK 220

**RESULT 14**

S59589 histone H1 - Chlamydomonas reinhardtii  
C/Species: Chlamydomonas reinhardtii  
C/Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 23-Jul-1999  
C/Accession: S59589; S62122  
R/Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, C.  
Curr. Genet. 28, 333-345, 1995  
A/Title: The organization structure and regulatory elements of Chlamydomona  
A/Reference number: S59581; MUID:96120862; PMID:8590479  
A/Accession: S59589  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-231 <FAB>  
C/Cross-references: EMBL:U16726  
A/Note: the authors did not translate the codon for residue 1  
R/Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
submitted to the EMBL Data Library, October 1994  
A/Description: The organization, structure and controlling elements of Chla  
A/Reference number: S62122  
A/Accession: S62122  
A/Molecule type: DNA  
A/Residues: 1-173, P', 174-231 <PAW>  
C/Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98452.1; PID:g571480  
C/Genetics:  
A/Introns: 62/3; 101/3  
C/Superfamily: histone H1  
C/Keywords: chromosomal protein; DNA binding; nucleosome

**Query Match** 31.7%; Score 164.5; DB 2; Length 231;  
**Best Local Similarity** 45.9%; Pred.No. 0.00038;  
**Matches** 56; Conservative 10; Mismatches 39; Indels 17; Gaps

Search completed: March 10, 2003, 12:28:30  
Job time : 32.9641 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 24.8354 Seconds  
(without alignments)  
182.035 Million cell updates/sec

Title: US-09-816-989A-7

Perfect score: 519

Sequence: 1 AKYAKAEKAYAKAKAAK.....AKYAKAEKAAKAEKAEAYEA 109

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
1	189.5	36.5	421	1 TOLA_ECOLI
2	181	34.9	210	1 H1_LYTP1
3	179	34.5	347	1 TOLA_PSEAE
4	175.5	33.8	344	1 MST1_DROHY
5	171.5	33.0	1391	1 MST2_DROHY
6	170.5	32.9	248	1 H1_PARAN
7	169.5	32.7	211	1 H1E_STRPU
8	165	31.8	219	1 H1I0_CHICK
9	163.5	31.5	384	1 TMPB_TREPH
10	163	31.4	224	1 H1L1_CHICK
11	160.5	30.9	206	1 H1L1_CHICK
12	160.5	30.9	225	1 H15_HUMAN
13	160	30.8	208	1 H1E_WYCSM
14	159.5	30.7	218	1 H1E_CHICK
15	158	30.4	218	1 H14_HUMAN
16	157	30.3	771	1 CALD_CHICK
17	156.5	30.2	218	1 H101_CHICK
18	154.5	29.8	240	1 H12_VOLCA
19	153.5	29.6	372	1 TOLA_HAETN
20	152.5	29.4	223	1 H103_CHICK
21	151	29.1	1403	1 YDF3_SCHPO
22	150	28.9	211	1 H12_MOUSE
23	150	28.9	217	1 H1G_STRPU
24	148	28.5	139	1 ASR_KLEPN
25	148	28.5	217	1 H1_AANAPL
26	147.5	28.4	218	1 H14_MOUSE
27	147	28.3	207	1 H11_CAEEL
28	147	28.3	220	1 H13_MOUSE
29	146	28.1	217	1 H1_CHICK
30	146	28.1	299	1 RL22_DROME
31	144	27.7	232	1 H12_GLYBA
32	144	27.7	352	1 ALGP_PSEAE
33	143.5	27.6	218	1 H12_RAT

34	143	27.6	214	1 DBH_MYCTU
35	142.5	27.5	221	1 H1C_CHITE
36	142	27.4	194	1 H1_SALTR
37	141.5	27.3	232	1 H1B_CHITE
38	141.5	27.3	345	1 RS6_AEDAE
39	140	27.0	232	1 H1A_CHITE
40	139	26.8	213	1 H13_RABIT
41	139	26.8	226	1 BASP_HUMAN
42	139	26.8	260	1 H11_VOLCA
43	138.5	26.7	1020	1 NPH_HUMAN
44	138	26.6	235	1 H1E_CHIPA
45	138	26.6	237	1 H1E_CHITE

## ALIGNMENTS

RESULT 1  
TOLA\_ECOLI  
ID TOLA\_ECOLI STANDARD; PRT; 421 AA.  
AC P19934;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE TOLA protein.  
GN TOLA OR CIM OR EXCC OR LKY OR B0739.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JMI05;  
RX MEDLINE=90078104; PubMed=2687247;  
RA Levengood S.K., Webster R.E.;  
RT "Nucleotide sequences of the tola and tolB genes and localization of  
their products, components of a multistep translocation system in  
Escherichia coli.";  
RL J. Bacteriol. 171:6600-6609(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP DOMAINS.  
RX MEDLINE=91296736; PubMed=2068069;  
RA Levengood S.K., Beyer W.F. Jr., Webster R.E.;  
RT "TolA: a membrane protein involved in colicin uptake contains an  
extended helical region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
RN [5]  
RP INTERACTION WITH PORINS.  
RX MEDLINE=97133271; PubMed=8978668;  
RA Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,

```

RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RL Lytechinus pictus and Strongylocentrotus purpuratus.";
RL Nucleic Acids Res. 14:8121-8133(1986).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04488; CAA238177.1; -.
DR PIR; A25550; A25550.
DR HSPSP; P02259; 1HST.
DR InterPro; IPR001386; Histone_H1/H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker_histone; 1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR KW . Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;

Query Match 34.9%; Score 181; DB 1; Length 210;
Best Local Similarity 50.0%; Pred. No. 1.3e-05;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

QY 2 KYAKKAEKAYAKAAKAEKKAYAKKAYAKAE-----AKKKAKEAKKAYAKAEAAK 57
DB 96 KTEAQKA-RAAAKAKALAAKKQKQKKAAKTKTKARKEKLAAKKAAKAAKVKKPAKAAK 154

QY 58 KEAYKAEAKKYAKAAKAEKKEKAAAAEAKKA-----EAAKAYKAEAAKAAAEKAA 106
DB 155 KPAAKAAKPPAAK--AAAKKPAAPKAAKPAKAAKPAKAAKPAKAAKPAKAAK 208

RESULT 3
TOLA PSEAE
ID TOLA_PSEAE STANDARD; PRT; 347 AA.
AC P50670;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
DE TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97113525; PubMed=8955385;
RT Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT "Identification and characterization of the tolQRA genes of
RT Pseudomonas aeruginosa";
RT J. Bacteriol. 178:7059-7068(1996).
RN [2]
RN REVISIONS TO N-TERMINUS.
RP Duan K., Sokol P.A.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

```













```
DR EMBL; M17020; AAA48790.1; -.
DR PIR; C28456; C28456.
DR HSP; P08287; IGH.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
FT INIT MET 0
FT DOMAIN 37 110 GLOBULAR.
SQ SEQUENCE 218 AA; 21672 MW; CB9724BFF14654A6 CRC64;

Query Match 30.7%; Score 159.5; DB 1; Length 218;
Best Local Similarity 47.7%; Pred. No. 0.0003;
Matches 52; Conservative 11; Mismatches 41; Indels 5; Gaps 3;

QY 1 AKYAKKAEKAYAKKAAKKEKAYAKK---EAKYAAEAKKAKAEAKKAYAKKAEAKAK 57
: || : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 SKKPGEGLEKAPKKKASAAKPKKAAAKKPAAPAAAKPKKAVAVKSPKKAKKPAASATKKS 168

QY 58 KEAYKAEAK-KYAKAAKAEKKEYAA-AEAKKAEAKKAYKAEAAKAAKE 104
:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 VKSPKAAKPKKAVAKSPAKAKAVKPKAAKPKAAKPKAAKAAKAAK 217

RESULT 15
H14_HUMAN
ID H14_HUMAN STANDARD; PRT; 218 AA.
AC P10412;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1.4 (Histone H1b).
GN H1F4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=87057092; PubMed=3782055;
RA Ohe Y., Hayashi H., Iwai K.;
RT "Human spleen histone H1. Isolation and amino acid sequence of a main
RT variant, H1b."
RL J. Biochem. 100:359-368(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9200931; PubMed=1916825;
RA Albig W., Kardalinos E., Drabant B., Zimmer A., Doenecke D.;
RT "Isolation and characterization of two human H1 histone genes within
RT clusters of core histone genes."
RL Genomics 10:940-948(1991).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: THIS VARIANT ACCOUNTS FOR 60% OF HISTONE H1.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M60748; AAA63187.1; -.
DR PIR; A24413; HSHU1B.
DR PIR; C40335; C40335.
DR HSP; P08287; IGH.
DR Genew; HGNC:4718; H1F4.
DR MIN; 142220; -.
```

```
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
KW Acetylation; Methylation.
FT INIT MET 0
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 25 25 METHYLATION (PARTIAL).
FT DOMAIN 35 113 GLOBULAR.
SQ SEQUENCE 218 AA; 21734 MW; E210BEC152ABE982 CRC64;

Query Match 30.4%; Score 158; DB 1; Length 218;
Best Local Similarity 48.6%; Pred. No. 0.00038;
Matches 53; Conservative 7; Mismatches 43; Indels 6; Gaps 3;

QY 2 KKYAKKAEKAYAKKAAKKEK---AYAKKAEKAYKAAEAKKAKAEAKKAEAKKAEAKKE 59
|| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 KKAASGEAKPKAKKAGAKAKKPKKATGAATPKKSAKTTPKKAKCPAAAA--G 165

QY 60 AYKAEAKKYAKKAAKAEK--KEYAAAAEAKKAEAKKAYKAEAAKAAKEAA 106
|||:: ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| |||
Db 166 AKKAKSPKAKAAKPKKAKPKSPAKAKAVKPKAAKPKTAKPKAAKPKKAA 214
```

Search completed: March 10, 2003, 12:17:11  
Job time : 25.8354 secs







```
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Membrane spanning protein, required for outer membrane integrity
DE TOLA OR Z0907 OR ECS0774.
GN Escherichia coli O157:H7.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AB005252; AAG55075.1; -.
DR EMBL; AP002553; BAB34197.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZEI.
KW Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5858D8E8230BDE28 CRC64;

Query Match 35.0%; Score 181.5; DB 16; Length 394;
Best Local Similarity 52.2%; Pred. No. 4.5e-05;
Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;

Oy 1 AKKYAKKAE-----KAYAKKAAKKKAYAKKEKAYKAAEAKKKAKAEKKYAK-EAAK 55
Db 120 ABEAAKQAEKQKQAEAAKAAADAKAKAEADDKA--AEAAKKAADAKKKAEAAK 177

Oy 56 AKKEAY-TAEAKKYAKAKAEKKYAAAEAKK--AEAAKAYKAAEAAKAAKEAA 106
Db 178 AAEEAQKAAEAAALKKKAAEAAEAAAEAAEAAEAAEAAEAAEAAEAAEAA 232

RESULT 6
ID Q828C1 PRELIMINARY; PRT; 376 AA.
AC Q828C1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE TOLA protein.
GN STY0793.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
```

```
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL627288; CAD05209.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZEI.
DR PRINTS; PR01574; TUBBYPROTEIN.
KW Complete proteome.
SQ SEQUENCE 376 AA; 38804 MW; EC21F2C4767A8A42 CRC64;

Query Match 34.3%; Score 178; DB 16; Length 376;
Best Local Similarity 52.3%; Pred. No. 7.4e-05;
Matches 58; Conservative 10; Mismatches 33; Indels 10; Gaps 4;

Oy 1 AKKYAKKAEKAYAKKAEKKYAKKEKAYKAAEAKKKAKAEKKYAKKEAAKKEA 60
Db 137 AKAADADAKKKAEAAEAAKAAADAKKKAEAEAA-AKAAADAKKKAEAEAA--AKAAAEAKK-- 190

Oy 61 YKAEAKKYAKAKAEKKYAAAEAKKAEAAKAYKAAEAAKAAAEAAEAAEAAE 108
Db 191 -KAEAAKAAADAKKKADAEAAKAAEAAKKAADAAKAAADAKKKAAAE 240

RESULT 7
ID Q8ZGZ2 PRELIMINARY; PRT; 388 AA.
AC Q8ZGZ2;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE TOLA colicin import membrane protein.
GN TOLA OR YP01123.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Kariyev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ414146; CAC89966.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZEI.
KW Complete proteome.
SQ SEQUENCE 388 AA; 40424 MW; 81447B04B30A7E7C CRC64;

Query Match 34.3%; Score 178; DB 16; Length 388;
Best Local Similarity 52.1%; Pred. No. 7.6e-05;
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

Oy 2 KKYAKKAEKAYAKKAEKKK-----AYAKKEA-KAYKA-AEAKKKAEKKYAKKEA 53
Db 140 KQAEQKQKIAAAAVAKAEKEEQKQAEATAAQAKAEADKIVKQAEAKKAEAAKAEAAVA 199

Oy 54 AKAKEAYKAEAKKYAKAAK-----AEKKEAYAAAEAKKAEAAK--AYKAEAAKAAKKEA 105
Db 200 AAANKQA-DADAKKAVEAEKAAADAEEKKAADAE-KKAAAKKVAEAAAEAKKAAAE 257

Oy 106 A 106
Db 258 A 258
```

```
DR HSP; PI9934; 1TOL.
DR InterPro; IPR00533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Complete proteome.
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;

Query Match 33.4%; Score 173.5; DB 16; Length 389;
Best Local Similarity 48.7%; Pred. No. 0.00015;
Matches 55; Conservative 18; Mismatches 31; Indels 9; Gaps 4;

QY 6 KKAEEKAYAKK-AKAAEK-KAYAKKEAKYAKAAEAKKKAKA--EAKYAKA-----AKA 56
DQ 145 KQAEAEAKQLAEAAKLAKEAEAKRLAALAKQAEAEAKKAEAEAKKAEAEAKKAEAEAKA 204
QY 57 KKEAYKAEAKYAKAAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 109
DQ 205 KVEKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 257

RESULT 10
O61164 PRELIMINARY; PRT; 1701 AA.
AC O61164;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM.
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235 (1998).
DR EMBL; AF031886; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;

Query Match 32.9%; Score 170.5; DB 5; Length 1701;
Best Local Similarity 53.5%; Pred. No. 0.00091;
Matches 54; Conservative 11; Mismatches 31; Indels 5; Gaps 5;

QY 5 AKKAEKAYAKKAAEKKAYAKKEAKYAKAAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 63
DQ 1205 AKKAEK-RKAEAVK-KAEAEAKKAEAKKAEAEKK-KAEAEAKKAEAEKKAEAEAKKAL 1261
QY 64 EAKYAKAAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 104
DQ 1262 ERKKAEAEAKKAEAEKKAEAEAKKAEAEKK-KAEAEAKKAEAE 1301

RESULT 11
Q39576 PRELIMINARY; PRT; 232 AA.
AC Q39576;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Histone H1.
GN CHI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
```

```
DR HSP; PI9934; 1TOL.
DR InterPro; IPR00533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
KW Complete proteome.
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;

Query Match 33.4%; Score 173.5; DB 16; Length 389;
Best Local Similarity 48.7%; Pred. No. 0.00015;
Matches 55; Conservative 18; Mismatches 31; Indels 9; Gaps 4;

QY 6 KKAEEKAYAKK-AKAAEK-KAYAKKEAKYAKAAEAKKKAKA--EAKYAKA-----AKA 56
DQ 145 KQAEAEAKQLAEAAKLAKEAEAKRLAALAKQAEAEAKKAEAEAKKAEAEAKKAEAEAKA 204
QY 57 KKEAYKAEAKYAKAAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 109
DQ 205 KVEKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 257

RESULT 10
O61164 PRELIMINARY; PRT; 1701 AA.
AC O61164;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM.
RX MEDLINE=98115903; PubMed=9448314;
RA Kappe S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235 (1998).
DR EMBL; AF031886; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87CE8A CRC64;

Query Match 32.9%; Score 170.5; DB 5; Length 1701;
Best Local Similarity 53.5%; Pred. No. 0.00091;
Matches 54; Conservative 11; Mismatches 31; Indels 5; Gaps 5;

QY 5 AKKAEKAYAKKAAEKKAYAKKEAKYAKAAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 63
DQ 1205 AKKAEK-RKAEAVK-KAEAEAKKAEAEAKKAEAEKK-KAEAEAKKAEAEKKAEAEAKKAL 1261
QY 64 EAKYAKAAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAEAKKAEAE 104
DQ 1262 ERKKAEAEAKKAEAEKKAEAEAKKAEAEKK-KAEAEAKKAEAE 1301

RESULT 11
Q39576 PRELIMINARY; PRT; 232 AA.
AC Q39576;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Histone H1.
GN CHI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Muller K., Lindauer A., Park P.B., Cornelius T., Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
```

```

RT histone genes reveal features linking plant and animal genes.";
RL Curr. Genet. 28:333-345(1995).
DR EMBL; U16726; AAA98452.1; -.
DR HSP; P02259; IHST.
DR InterPro; IPR001386; Histone H1/H5.
DR InterPro; IPR003216; Linkerhist.N.
DR Pfam; PF00538; linker histone; I.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
SQ SEQUENCE 232 AA; 24693 MW; 2D006AE44A8FA037 CRC64;

Query Match 32.8%; Score 170; DB 10; Length 232;
Best Local Similarity 46.3%; Pred. No. 0.00016;
Matches 57; Conservative 10; Mismatches 38; Indels 18; Gaps 5;

QY 2 KKYAKKA---EKAYAKKAAKAAEKKAYAKKEA---KAYKAAEKKKKAAKAAEKKYAKKA- 53
Db 100 KKAAKKAATPKAAAPKKEGAVKTKAPKKEGKKPKSAKKAEEKPKKEGKKAAKPAK 159

QY 54 ----AKAKKEAYKAEAKKYAKAAKAAEKKEKYAAAE---AKCAEAAK---AYKAAEAATAAK 103
Db 160 AEKKPKAAKPAKKTPTTKAAAKPKAEKKPKAAAKPKAEKKPKAAKPKAEKKAKAAKPAK 219

QY 104 EAA 106
:|

```

Query Match	32.1%	Score 166.5;	DB 5;	Length 619;
Best Local Similarity	41.0%;	Pred. No. 0.00067;		
Matches	55;	Conservative 16;	Mismatches 30;	Indels 33; Gaps 4;
Qy	9	EKAVAKKAAKKBKAVAKKEAKYKAAEAKKAKAKAEKKVAKKEAAK-AKKEAVKAE--A	65	
Db	331	EKKTAQKADAEKREERALKAEERDRLTAEAKQAAAKKAKAEAEAKIAAEALLAEAAA	390	
Qy	66	KKYKAAKAA-----EKKEEYAAAEAKCAEAAKAYKA--	95	
Db	391	QKAAEEAKALKAAEDAAQKAAEEARLAEAEAAAKQVAAEEAAQKAAEEARLAEAEAAAKAAE	450	
Qy	96	EAAKAAAEKAEAYEA	109	
Db	451	EAAQKAAEEEAALKAA	464	

RESULT 14	
Q39598	PRELIMINARY; PRT; 265 AA.
ID	
AC	Q39598;
DT	01-NOV-1996 (T-EMBLrel. 01, Created)
DT	01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE	Cgcr-4 product (Fragment).
GN	CGCR-4.
OS	Chlamydomonas reinhardtii.
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC	Chlamydomonadaceae; Chlamydomonas.
OX	NCBI_TaxID=3055;
RN	[1]
RP	SEQUENCE FROM N.A.

KA MEDLINE=921119224; PUBMED=1/31960;  
RA MAKARCHUK W.W., Muller F.W., Beck C.F.;  
RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex  
RT arrangements of directly repeated sequence motifs.";  
RL Plant Mol. Biol. 18:143-146(1992).  
DR EMBL; X17208; CAAJ5080.1; -.  
FT NON TER 1  
SQ SEQUENCE 265 AA; 26216 MW; B35318B7377CF782 CRC64;

Query Match 31.8%; Score 164; DB 10; Length 265;  
Best Local Similarity 39.4%; Pred. No. 0.00045;  
Matches 50; Conservative 21; Mismatches 38; Indels 18; Gaps 2;

Qy 1 AKYKAKAEKA-----YAKKAKAKKEKKA YAKKEAKAYKAAEAKKKAKAEA 46



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:01 ; Search time 56.7996 Seconds  
(without alignments)  
255.712 Million cell updates/sec

**Title:** US-09-816-989A-7

Perfect score:

Sequence: 1 AKKYAKKAEKAYAKKAKAAK.....AKAYKAEAAKAAKAAAYEA 109

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%  
Maximum Match 100%

Maximum Match 100%  
Listing first 45 summaries

Database : A Genesec 101002: \*

Database : A\_Geneset\_10102

1:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1980.DAT.*
2:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1981.DAT.*
3:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1982.DAT.*
4:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1983.DAT.*
5:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1984.DAT.*
6:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1985.DAT.*
7:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1986.DAT.*
8:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1987.DAT.*
9:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1988.DAT.*
10:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1989.DAT.*
11:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1990.DAT.*
12:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1991.DAT.*
13:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1992.DAT.*
14:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1993.DAT.*
15:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1994.DAT.*
16:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1995.DAT.*
17:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1996.DAT.*
18:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1997.DAT.*
19:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1998.DAT.*
20:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1999.DAT.*
21:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2000.DAT.*
22:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2001.DAT.*
23:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	519	100.0	109	21 AAY82577	Copolymer molecule
2	312.5	60.2	86	21 AAY82576	Copolymer molecule
3	289	55.7	77	21 AAY82575	Copolymer molecule
4	228.5	44.0	66	21 AAY82574	Copolymer molecule
5	209	40.3	154	11 AAR06445	Recombinant copoly
6	187	36.0	106	11 AAR06446	Recombinant copoly
7	180.5	34.8	56	21 AAY82573	Copolymer molecule
8	179.5	34.6	100	21 AAY98499	Peptide #10 used i
9	179.5	34.6	100	21 AAY59044	Amino acid polymer
10	179.5	34.6	100	22 AAU04289	Poly(Lys-Ala) used

11	179.5	34.6	100	22	AB45852	Nucleic acid trans
12	178	34.3	223	20	AA14928	Amino acid sequenc
13	162.5	31.3	643	16	AR84568	Trypanosoma cruzi
14	160.5	30.9	158	20	AA34068	Histone H1 isoform
15	160.5	30.9	158	21	AA34068	Human histone H1.5
16	160.5	30.9	226	20	AA34060	Human histone H1.1
17	160.5	30.9	226	21	AA34060	Human histone H1.1
18	159.5	30.7	234	23	AA57358	Histone H1-4, frag
19	159.5	30.7	234	23	AAU0944	Human linker hist
20	158.5	30.5	472	16	AR84569	Trypanosoma cruzi
21	158	30.4	218	18	AA29477	Human histone H1.1
22	158	30.4	218	20	AA34034	Human histone H1.1
23	158	30.4	218	21	AA57332	Human histone H1.1
24	157.5	30.3	564	16	AR84565	Trypanosoma cruzi
25	154.5	29.8	130	23	AAU0945	Peptide fragment o
26	154.5	29.8	130	23	AAE13237	Human linker hist
27	154	29.7	219	21	AA76981	Human linker hist
28	152.5	29.4	222	18	AA29476	Quail H1 histone p
29	152.5	29.4	222	20	AA34033	Human histone H1.1
30	152.5	29.4	222	21	AA57331	Human histone H1.1
31	151	29.1	607	22	AB62173	Human histone H1.1
32	150	28.9	1507	21	AA24128	Drosophila melanog
33	148.5	28.6	140	13	AB25206	Plasmodium yoelii
34	146	28.1	299	22	AB63276	Synthetic helical
35	144	27.7	111	18	AA44829	Drosophila melanog
36	144	27.7	434	18	AAW15291	Adhesive polypepti
37	144	27.7	434	21	AA56903	Antibiotic potenti
38	143	27.6	214	20	AA34055	(Lys)434 protein s
39	143	27.6	214	20	AA57353	M. tuberculosis hi
40	141.5	27.3	239	23	AB49123	M. tuberculosis hi
41	141	27.2	100	21	AA59849	Listeria monocytog
42	141	27.2	100	21	AA59042	Peptide #8 used in
43	141	27.2	100	21	AA59042	Amino acid polymer
44	141	27.2	100	22	AAU04287	Poly-L-Lysine used
45	141	27.2	100	22	AB45850	Nucleic acid trans
46	138.5	26.7	617	22	AB35445	Peptide #2951 enc

## ALIGNMENTS

## RESULT 1

RESOLUTION 1  
AA82577  
ID AAY82577 standard; peptide: 109 AA.

XX AAY82577:

AC XX  
AA102511;

28-JUL-2000 (first entry)

DE	20-000 2000 (11182 entry)	XX	DE	Copolymer molecular weight	TV-marker amino acid sequence	SEQ ID NO: 7
DE			DE			

Copolymer; molecular weight IV marker; amino acid sequence only.

XX KW Copolymer; molecular weight marker; TV-marker; immune disease;  
KW Glutramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;  
KW antidiabetic; thyromimetic; haemostatic; antiparastic; dermatological;  
KW antanaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KW Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
KW Hashimoto's disease; idiopathic myxoedema;  
KW pemphigus vulgaris; systemic lupus erythematosus.

Identified

US  
XX  
Unidentified.

XX PN WO200018794-

PN W0200018/94-A1.  
XX

PD 06-APR-2000.

XX  
73  
00-4444-0000.

PF 24-SEP-1999; 99WO-US22402





Db 26 -KABAKKYAKAAKAEKKYEAEEAK-----YKAEAAKAAKAEKAEYEA 66

## RESULT 5

AAR06445  
ID AAR06445 standard; protein; 154 AA.

XX AAR06445;

XX 03-JAN-1991 (first entry)

XX Recombinant copolymer 1-77, myelin basic protein analogue.

XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;

XX immunological activity; autoimmune encephalomyelitis;

XX multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX 22-AUG-1990.

XX 16-FEB-1990; 90EP-0301700.

XX 07-FEB-1990; 90US-0473845.

XX 17-FEB-1989; 89US-0312541.

XX (REPL-) REPLIGEN CORP.

XX Cook KS;

XX WPI; 1990-255848/34.

XX N-PSDB; AAQ05664.

XX Producing genes encoding random polymers of aminoacid(s) - for

XX producing recombinant polypeptide(s) with biological and/or

XX immunological activity

XX Disclosure; Fig 11; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes

XX coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaTAN

XX (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to

XX express Protein A. The resulting plasmids encode fusion proteins

XX consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.

XX A methionine residue occurs between the Protein A and rCOP-1

XX sequences, originating from the 5' linker sequence, in order that

XX the COP-1 polypeptide may be cleaved from the fusion protein.

XX rCOP-1-77 contains oligonucleotide duplexes encoding the following

XX segments: YKK, EAE, KAK, AAK, and AAA. The N-terminal alanine residue

XX is left behind following CNBr cleavage of the fusion protein.

XX The product prevents or arrests experimental autoimmune

XX encephalomyelitis. They are used to prevent, arrest or control a

XX demyelinating disorder, e.g. multiple sclerosis. They may also

XX be used as additives to hair care products to confer beneficial

XX effects on damaged hair or as supplements for diets deficient in

XX certain amino acids.

XX See also AAQ05665.

XX Sequence 154 AA;

SQ Query Match 40.3%; Score 209; DB 11; Length 154;

Best Local Similarity 53.7%; Pred. No. 1.6e-10;

Matches 72; Conservative 8; Mismatches 20; Indels 34; Gaps 9;

QY 2 KKYAKKAEKA-----YAKKAAKAEKKYEA-----KKEAKYKAEAA-----KKKAK-A 44

Db 24 KKKAKAEKAKKAKYKKYKKEAEAAKAAKAAKAAAYKKAEAA--AAEAEKAKYKKKAEKA 81

QY 45 EAKKYAEKAEKA--KKAEYKAEKAKYKAKAEK-----KEYAAAEKAEAA--K 91

Db 82 EYKCKAKAAAEAYKKAEAEYKKYKKKAKKAKYKKKAEKAKAAAEAEKAEAEYK 141

QY 92 AYKAEAAKAAKAEKA 105

Db 142 KYKKEAEK--AKEA 153

## RESULT 6

AAR06446

ID AAR06446 standard; protein; 106 AA.

XX AAR06446;

XX 03-JAN-1991 (first entry)

XX Recombinant copolymer 1-19, myelin basic protein analogue.

XX Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;

XX immunological activity; autoimmune encephalomyelitis;

XX multiple sclerosis;

XX Synthetic.

XX EP383620-A.

XX 22-AUG-1990.

XX 16-FEB-1990; 90EP-0301700.

XX 07-FEB-1990; 90US-0473845.

XX 17-FEB-1989; 89US-0312541.

XX (REPL-) REPLIGEN CORP.

XX Cook KS;

XX WPI; 1990-255848/34.

XX N-PSDB; AAQ06446.

XX Producing genes encoding random polymers of aminoacid(s) - for

XX producing recombinant polypeptide(s) with biological and/or

XX immunological activity

XX Disclosure; Fig 12; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes

XX coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaTAN

XX (deposit: 20-NOV-1984 US4691009, NRRL B-15910), a plasmid used to

XX express Protein A. The resulting plasmids encode fusion proteins

XX consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.

XX A methionine residue occurs between the Protein A and rCOP-1

XX sequences, originating from the 5' linker sequence, in order that

XX the COP-1 polypeptide may be cleaved from the fusion protein.

XX rCOP-1-19 contains oligonucleotide duplexes encoding the following

XX segments: YKK, AAE, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The

XX N-terminal alanine residue is left behind following CNBr cleavage of the

XX fusion protein.

XX The product prevents or arrests experimental autoimmune

XX encephalomyelitis. They are used to prevent, arrest or control a

XX demyelinating disorder, e.g. multiple sclerosis. They may also

XX be used as additives to hair care products to confer beneficial

XX effects on damaged hair or as supplements for diets deficient in

XX certain amino acids.

XX See also AAQ05664.

XX Sequence 106 AA;

SQ Query Match 36.0%; Score 187; DB 11; Length 106;

Best Local Similarity 55.0%; Pred. No. 7.7e-09;

Matches 60; Conservative 11; Mismatches 20; Indels 18; Gaps 6;

QY 5 AKKAEKAYAKKAAKKEKAYAKKEAKAYKAAAEKKAKEAKKAYKAEKAAKKEAYKAE 64

Db 1 AKAAEKAKAAK--KAYEAEKAKAKYBAK--KAEKAEKAEKAAAEKKAKEAKKAEK-- 51





Query Match	34.6%	Score 179.5	DB 22	Length 100
Best Local Similarity	55.6%	Pred. No. 3.1e-08		
Matches	55	Conservative	8	Mismatches 33
				Indels 3
				Gaps 3





```

RESULT 15
AAAY57366
ID AAAY57366 standard; peptide; 158 AA.
XX AC
XX AC AAAY57366;
XX XX
DT DT
XX XX
XX 13-JUN-2000 (first entry)
XX XX
DE Human histone H1.5 pANCA-reactive fragment (residues 69-226).
XX XX
XX Ulcerative colitis; inflammatory bowel disease; porin antigen;
KW pANCA; perinuclear anti-neutrophil cytoplasmic antibody; human;
KW histone H1; isoform; NANUC-2.
XX KW
OS Homo sapiens.
XX OS
PN US6033864-A.
XX PN
XX XX
XX 07-MAR-2000.
XX PD
XX XX
XX 12-MAR-1998; 98US-0041889.
XX PF
XX XX
XX 12-APR-1996; 96US-0057846.
XX PR
XX 11-APR-1997; 97US-0837058.
XX PR
XX XX
PA (REGC ) UNIV CALIFORNIA.
XX PA
XX XX
XX Cohavy O, Braun J;
XX PI
XX XX
DR WPI; 2000-255695/22..
XX DR
XX XX
PT Diagnosing ulcerative colitis or susceptibility, by detecting complex
PT formation between microbial porin antigen and perinuclear
PT anti-neutrophil cytoplasmic autoantibodies -
XX PT
XX XX
XX Example 4; Columns 69-70; 49pp; English.
XX PS
XX XX
CC The invention provides a method for diagnosing ulcerative colitis in a
CC subject suspected of having inflammatory bowel disease. The method
CC comprises reacting a patient sample with a porin antigen that is
CC immunologically reactive with pANCA (perinuclear anti-neutrophil
CC cytoplasmic antibodies) and detecting formation of a Ag-pANCA complex
CC as indicative of ulcerative colitis. The method is used to diagnose
CC ulcerative colitis or susceptibility to it. Sequences AAY57362-367
CC represent pANCA-reactive fragments, derived from human histone H1.5.
XX CC
SQ Sequence 158 AA;

Query Match 30.9%; Score 160.5; DB 21; Length 158;
Best Local Similarity 46.8; Pred. No. 1.9e-06;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

QY 1 AKKYAKKAEKAVAKAKAEKKYAVKKEAKYKAA-----EAKKKAKAEAKKYVKAERAK 55
DB 51 AKPKAKKGAGAKAKKPAGATPK--AKKAAGAKKAVKTTPKKAKKPAAAGVKVKAESPCK 108
QY 56 AKCEYAKAEAKKYAKAKAEKKEYAAAAEKAEAAKAYKABAAAEEAKE 104
DB 109 AKAAAKPKKATKSPAKPKAVPKAHPKAAPKPAKPAKPAKPAKPAKPAKPAK 157

Search completed: March 10, 2003, 12:21:31
Job time : 57.7996 secs

```

**THIS PAGE BLANK (USPTO)**

Result No.	Query			Description	
	Score	Match	Length	ID	Description
1	519	100.0	109	10	US-09-816-989A-7
2	312.5	60.2	86	10	US-09-816-989A-6
3	289	55.7	77	10	US-09-816-989A-5
4	228.5	44.0	66	10	US-09-816-989A-4
5	180.5	34.8	56	10	US-09-816-989A-3
6	178	34.3	223	9	US-10-051-643-201
7	162	31.2	356	9	US-09-820-843A-27
8	154.5	29.8	309	9	US-09-820-843A-24
9	153.5	29.6	369	9	US-09-820-843A-95
10	153.5	29.6	372	9	US-09-820-843A-8
11	144	27.7	352	9	US-09-820-843A-23
12	144	27.7	434	10	US-09-124-280A-3
13	143.5	27.6	452	9	US-10-184-832-5
14	138.5	26.7	617	10	US-09-964-761-36182
15	138	26.6	45	10	US-09-816-989A-2
16	126	24.3	600	9	US-09-738-626-5197
17	122.5	23.6	218	9	US-09-999-724-48
18	122.5	23.6	892	10	US-09-815-242-11216
19	122.5	23.6	892	10	US-09-815-242-13765
20	122.5	23.6	892	10	US-09-815-242-13765

US-09-816-985A-6  
: Sequence 6. Application US/09816989A

```
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match      60.2%; Score 312.5; DB 10; Length 86;
Best Local Similarity 72.1%; Pred. No. 1.5e-18;
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5;

QY 1 AKKYAKAEKAYAKAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAE 60
DB 1 AKKYAKK-EKAYAKKA-----EKAAKAEKAYAKAEKKAKKA----- 36

QY 61 YKAEAKKYAKAKAEKKKEYAAAEAK-KAEAA-KAYKAEAAKAAKAEAAAYEA 109
DB 37 -KAEAKKYAKAKAEKKKEYAAAEAKYKAEAAKKAYKAEAAKAEAAAYEA 86

RESULT 3
US-09-816-989A-5
; Sequence 5, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match      55.7%; Score 289; DB 10; Length 77;
Best Local Similarity 67.0%; Pred. No. 9.1e-17;
Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4;

QY 1 AKKYAKAEKAYAKAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAE 60
DB 1 AKKYAKK-EKAYAKKA-----EKAAKAEKAYAKAEKKAKKA----- 36

QY 61 YKAEAKKYAKAKAEKKKEYAAAEAK-KAEAA-KAYKAEAAKAAKAEAAAYEA 109
DB 37 -KAEAKKYAKAKAEKKKEYAAAEAKYKAEAAKKAYKAEAAKAEAAAYEA 77
```

## RESULT 4

```
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4
```

```
Query Match      44.0%; Score 228.5; DB 10; Length 66;
Best Local Similarity 56.9%; Pred. No. 4.2e-12;
Matches 62; Conservative 0; Mismatches 4; Indels 43; Gaps 4;

QY 1 AKKYAKAEKAYAKAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAE 60
DB 1 AKKYAKK-EKAYAKKA-----EKAAKKA----- 25

QY 61 YKAEAKKYAKAKAEKKKEYAAAEAK-KAEAA-KAYKAEAAKAAKAEAAAYEA 109
DB 26 -KAEAKKYAKAKAEKKKEYAAAEAK-----YKAEAAKAAKAEAAAYEA 66
```

## RESULT 5

```
US-09-816-989A-3
; Sequence 3, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3
```

```
Query Match      34.8%; Score 180.5; DB 10; Length 56;
Best Local Similarity 45.9%; Pred. No. 2e-08;
Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3;

QY 1 AKKYAKAEKAYAKAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAEKKAYAKAE 60
DB 1 AKKYAKK-EKAYAKKA-----YKAEAAKAAKAEAAAYEA 56
```



Db 1 AKYAKK-EKAYAKK-----EKAAKKAEKAYKAAEAKKKAEAK-----39  
QY 61 YKAEAKKYAKAAKAEKKYAAAEAKKAEAAKAYKABAAKAAKAAEAYEA 109  
Db 40 -----YKAEAAKAAKAEAYEA 56

## RESULT 6

US-10-051-643-201  
; Sequence 201, Application US/10051643  
; Publication No. US20020197265A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Tan, Paul L. J.  
; TITLE OF INVENTION: Methods and Compounds for the Treatment  
; TITLE OF INVENTION: Of Immunologically-Mediated Diseases of the Respiratory  
; TITLE OF INVENTION: System using Mycobacterium Vaccae  
; FILE REFERENCE: 11000.1008c2  
; CURRENT APPLICATION NUMBER: US/10/051,643  
; CURRENT FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US09/156,181  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: US 08/996,624  
; PRIOR FILING DATE: 1997-12-23  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 201  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
US-10-051-643-201

Query Match 34.3%; Score 178; DB 9; Length 223;  
Best Local Similarity 54.5%; Pred. No. 1.3e-07;  
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;  
QY 1 AKYAKK-AEKYAKKAAKAE-----KKAYAKKEKAYKAAEAKKKAKAEAKKYAKAEAK 55  
Db 112 ARKAAPAKPAKAAKAAKAPAKPAKKA-ATKAAPAKKATAAKKAAKAPAKKATAAKKAAAP 170  
QY 56 AKKEAYKAEAKKYAKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAK 103  
Db 171 AKKAPAKKAATKAAKAPAKKAPAKKAAATKAAKAPAKKAPAAK--KAPAKKAPAK 218

## RESULT 7

US-09-820-843A-27  
; Sequence 27, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 27  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Vibrio cholerae  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: tola protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|9656364  
US-09-820-843A-27

Query Match 31.2%; Score 162; DB 9; Length 356;  
Best Local Similarity 42.6%; Pred. No. 3.6e-06;  
Matches 55; Conservative 13; Mismatches 37; Indels 24; Gaps 5;

QY 3 KYAKKAEKAYAKK--AKAAEKKAYAKKEAKKAYKAAEAKKK-----AKAEAKKYAKAEAKA 56  
Db 112 KAAREAEKLREQEQERLAAEQKAREEKERAAKAEAKKVKBEAAKKAQERVAKEAAAA 171  
QY 57 KKEAYKAEAKKYAK-----AKAEK-----KEVAAAEAKKAEAAKAYKAEAKA 100  
Db 172 KAEQORIEREKAKLAEBEKAKREKEVAAKAEQERLAKERAKEA--ADKAKKEKERAKA 229  
QY 101 AAKEAAVEA 109  
Db 230 EAERKAQEA 238

## RESULT 8

US-09-820-843A-24  
; Sequence 24, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 24  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: polyhydroxyalkanoate synthesis protein Phaf  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|9951352  
US-09-820-843A-24

Query Match 29.8%; Score 154.5; DB 9; Length 309;  
Best Local Similarity 46.9%; Pred. No. 1.2e-05;  
Matches 53; Conservative 3; Mismatches 50; Indels 7; Gaps 2;  
QY 1 AKKYAKKAEKAYAKKAAKAEKKYAKKAEK-----AYKAAEAKKKAKAEAKKYAKAEAA 54  
Db 150 AKPAKPAKTAATAAKPAKPAKAAKAAKPAKPAKKTAAKTAATAAKPAKPAKPTAKAAA 209  
QY 55 K-AKKEAYKAEAKKYAKKAEKKEYAAEAKKAEAAKAYKAEAAKAAKAEAA 106  
Db 210 KPATKPAKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAATAAKPAKPAKPAKPKPA 262

## RESULT 9

US-09-820-843A-95  
; Sequence 95, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: T. pallidum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: conserved hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|3322751  
US-09-820-843A-95



```

/ PRIOR APPLICATION NUMBER: PCT/US01/00682
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 36182
/ LENGTH: 617
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005529.7
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
/ OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUE 2.00e-33
/ OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUE 3.00e-22
US-09-864-761-36182

Query Match          26.7%; Score 138.5; DB 10; Length 617;
Best Local Similarity 35.3%; Pred.No. 0.00044;
Matches 49; Conservative 17; Mismatches 36; Indels 37; Gaps

QY 5 AKKAEKAYAK---KAKAAKEKKAYAKKEAYKAAEAKKAKAEAKKYAKEAKAK---- 57
    ||| ||| : : : : : ||| : : : ||| : : : ||| : : : |||
Db 161 AKSPEKEAKSPAENVKSPKSAKSPKAEKEAKSPAEEKSPKSAKSPKAEAKSPVK 220
    ||| ||| : : : : : ||| : : : ||| : : : ||| : : : |||

QY 58 -----KEAYKAAKKYAKAAKAEKKEAYAA-----AEAKKAEAKA- 92
    ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 221 EBAKSPAENVKSPKSAKSPKKEAKSPKSAKSPKAEKEAKSPAEEKSPKSAKSPK 280
    ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 93 YKAEA-----AKAAAKEAA 106
    ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| |||

Db 281 VKAEAKSPAENVKSPKKEA 239
    ||| ||| : : : ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-816-989A-2
/ Sequence 2, Application US/09816989A
/ Patent No. US20020115103A1
/ GENERAL INFORMATION:
/ APPLICANT: Lis, Doris
/ TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECUL
/ FILE REFERENCE: 2609/60807-A-PCT-US
/ CURRENT APPLICATION NUMBER: US/09/816,989A
/ CURRENT FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: 60/101,693
/ PRIOR FILING DATE: 1998-09-25
/ PRIOR APPLICATION NUMBER: PCT/US99/22402
/ PRIOR FILING DATE: 1999-09-24
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 2
/ LENGTH: 45
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptid
US-09-816-989A-2

```

Qy 65 AKKYAKAAKAE--KKEYAAAEAKKAAEAAKAYKAAEAAKAAKAAEAAEYEA 109  
Db 1 AKKYAKKAAEAKKAKKAYKAAEAKK--AAKYEKAAAEKAAKAAEAAEYEA 45

Search completed: March 10, 2003, 12:53:45  
Job time : 33.5669 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2003, 12:15:04 ; Search time 19.3165 Seconds  
(without alignments)  
166.029 Million cell updates/sec

Title: US-09-816-989A-7  
Perfect score: 519  
Sequence: 1 AKYAKAEKAYAKAKAAK.....AKYAKAEKAAKAAKAAVEA 109

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179.5	34.6	100	2	US-08-460-890A-64
2	179.5	34.6	100	3	US-08-167-641C-64
3	179.5	34.6	100	4	US-08-460-971A-64
4	179.5	34.6	100	4	US-08-462-040-64
5	178	34.3	223	4	US-09-095-855-201
6	178	34.3	223	4	US-09-205-426-201
7	162.5	31.3	643	2	US-08-216-894-8
8	162.5	31.3	643	4	US-09-115-746-8
9	160.5	30.9	158	3	US-09-041-889-40
10	160.5	30.9	226	3	US-09-041-889-32
11	158.5	30.5	472	2	US-08-216-894-10
12	158.5	30.5	472	4	US-09-115-746-10
13	158.5	30.5	564	2	US-08-216-894-2
14	158.5	30.5	564	4	US-09-115-746-2
15	158	30.4	218	3	US-09-041-889-4
16	158	30.4	218	3	US-08-837-058-3
17	152.5	29.4	222	3	US-09-041-889-3
18	152.5	29.4	222	3	US-08-837-058-3
19	150	28.9	1507	3	US-08-929-329-5
20	144	27.7	434	1	US-08-097-830E-3
21	144	27.7	434	2	US-08-456-112B-3
22	143	27.6	214	3	US-09-041-889-27
23	141	27.2	100	2	US-08-460-890A-62
24	141	27.2	100	3	US-08-167-641C-62
25	141	27.2	100	4	US-08-460-971A-62
26	141	27.2	100	4	US-08-462-040-62
27	133.5	25.7	212	3	US-09-041-889-1

28	133.5	25.7	212	3	US-08-837-058-1
29	133	25.6	433	1	US-08-346-849-2
30	133	25.6	433	2	US-08-293-284A-2
31	129.5	25.0	113	3	US-09-041-889-14
32	129.5	25.0	113	3	US-08-837-058-14
33	129.5	25.0	147	3	US-09-041-889-13
34	129.5	25.0	147	3	US-08-837-058-13
35	129.5	25.0	220	3	US-09-041-889-2
36	129.5	25.0	220	3	US-08-837-058-2
37	128	24.7	262	1	US-08-403-379A-1
38	128	24.7	262	2	US-08-929-414-1
39	128	24.7	263	2	US-08-557-309B-51
40	128	24.7	263	3	US-08-834-306-51
41	128	24.7	263	4	US-08-993-674A-51
42	128	24.7	263	4	US-09-256-976-51
43	126.5	24.4	56	4	US-08-993-008A-6
44	124.5	24.0	1561	3	US-08-894-017-23
45	123.5	23.8	516	2	US-08-762-106-8

ALIGNMENTS

RESULT 1  
US-08-460-890A-64  
; Sequence 64, Application US/08460890A  
; Patent No. 5994109  
; GENERAL INFORMATION:  
; APPLICANT: WOO Savio L.C.  
; APPLICANT: Smith, Louis C.  
; APPLICANT: Cristiano, Richard J.  
; APPLICANT: Gottchalk, Stephen  
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND  
; TITLE OF INVENTION: METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,890A  
FILING DATE: June 5, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/167,641  
FILING DATE: December 14, 1993  
APPLICATION NUMBER: 07/855,389  
FILING DATE: March 20, 1992  
APPLICATION NUMBER: PCT/US93/02725  
FILING DATE: March 19, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 212/066  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

Sequence 1, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 13, Appl  
Sequence 13, Appl  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 51, Appl  
Sequence 51, Appl  
Sequence 51, Appl  
Sequence 6, Appli  
Sequence 23, Appl  
Sequence 8, Appli





```
RESULT 6
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-205-426-201

Query Match      34.3%; Score 178; DB 4; Length 223;
Best Local Similarity 54.5%; Pred. No. 2.2e-08;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

Qy 1 AKYAKK--AKYAKKAKAKE---KKYAKKEAKYKAAEAKKKAKAEKKAAYKAEAK 55
Db 112 ARKAAPKAPAKKAAKAAKAPAKKAPAKKA-ATKAAPKAKTAATAKGAAPAKKATAAKKAA 170
Qy 56 AKCEAYKAEAKYKAAKAEKKEYA--AAEAKKAEAAKAYKAEAAKAAK 103
Db 171 AKKAPAKKAAATKAPAKKAPAKKAPAKKATKAPAKKAPAKK--KAPAKKAPAK 218

RESULT 7
US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8

Query Match      31.3%; Score 162.5; DB 4; Length 643;
Best Local Similarity 44.4%; Pred. No. 1.3e-06;
Matches 52; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

Qy 1 AKKYAKKAEKAYKAKAEKKEKAYKAEKAEKKAAYKAEKKAKEAKKAEAKKYA---KEA 56
Db 445 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 504
Qy 57 KK--EAYKAEAKKYKAAKAEKKEYA----AAEAKKAEAAKAYK--AEAAKAAKAEAA 106
Db 505 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 561

RESULT 8
US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-8

Query Match      31.3%; Score 162.5; DB 4; Length 643;
Best Local Similarity 44.4%; Pred. No. 1.3e-06;
Matches 52; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

Qy 1 AKKYAKKAEKAYKAKAEKKEKAYKAEKAEKKAAYKAEKKAKEAKKAEAKKYA---KEA 56
Db 445 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 504
```



```

QY 57 KK--EAYVAEAKKYAKAAAEKKEYYA-----AAEAKKAAEAAKAYK-ABAAKAAAEAA 106
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 505 TKVAEAEKQKAAEATKVAAEAKQKAAEATKVAAEAEKQKAAEATKVAAEAEKQKAAEAA 561

RESULT 9
US-09-041-889-40
; Sequence 40, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/837,058
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-041-889-40

Query Match 30.9%; Score 160.5; DB 3; Length 158;
Best Local Similarity 46.8%; Pred. No. 4.6e-07;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

QY 1 AKKYAKKAEAYAKKAAAEKKEKAYAKKEAYKAA-----EAKKKAAEAKKYAKEAAK 55
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 51 AKPKAKKAGAAKAKKPAGATPKK--AKKAAGAKVAVKTKPKAKKPPAAAGVKVAKSPKK 108
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 56 AKKEAYKAEAKKYAKAAAEKKEYYAABAAEAKKAAEAAKAYKAEAAKAAAE 104
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 109 AKAAAKPKKATKSPAKPKVAVKPKAAKPKAAKPKAAKPKAAKPKAAKPKAAKPKAAK 157
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
US-09-041-889-32
; Sequence 32, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens

```

Matches 51; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 1 AKKYAKAEKAYAKAKAEKAYAKAEKAYAKAEKAKKAKAEAKKYA----KEAAXA 56  
Db 315 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 374  
QY 57 KK--EAYKAEKKYAKAAKAEKKEYA----AAEKKAEEAAKAYK-AAEAKAAAE 105  
Db 375 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 430

RESULT 13  
US-08-216-894-2  
; Sequence 2, Application US/08216894  
; Patent No. 5876734  
; GENERAL INFORMATION:  
; APPLICANT: Kirchoff, Louis V.  
; APPLICANT: Otsu, Keiko  
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/216,894  
; FILING DATE: 24-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 564 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-216-894-2

Query Match 30.5%; Score 158.5; DB 2; Length 564;  
Best Local Similarity 44.0%; Pred. No. 2.4e-06;  
Matches 51; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 1 AKKYAKAEKAYAKAKAEKAYAKAEKAYAKAEKAKKAKAEAKKYA----KEAAXA 56  
Db 431 AAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 490  
QY 57 KK--EAYKAEKKYAKAAKAEKKEYA----AAEKKAEEAAKAYK-AAEAKAAAE 105  
Db 491 TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEA 546

RESULT 14  
US-09-115-746-2  
; Sequence 2, Application US/09115746  
; Patent No. 6228601  
; GENERAL INFORMATION:  
; APPLICANT: Kirchoff, Louis V.  
; APPLICANT: Otsu, Keiko  
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI



THIS PAGE BLANK (USPTO)